اللسانيات الحاسوبية

تنظيراً وتطبيقاً

الاستاذ المساعد الدكتور
خالد حوير الشمس





أ. م. د. خالد حوير الشمس

اللسانيات الحاسوبية تنظيراً وتطبيقاً



تقديم

قطعت اللسانيّات في العالم أشواطًا من التطوّر والتجديد حتى بلغت مرحلة صارت فيها من العلوم المتطوّرة بتأثير العلاقة الجدليّة بين اللسانيّات، والعلوم الأخرى، وتحوّل الدرس اللغويّ في أوربا من دراسة شكليّة إلى درس علميّ يُعنى بدراسة اللغات العالميّة دراسة علميّة؛ وبذلك صارت آراء دي سوسير انتقالة نوعيّة فتحت الآفاق لتطوّر الدرس اللسانيّ ومكّنته من أن يكون مثار اهتمام علماء النفس، والفلاسفة، وعلماء الاجتماع، الذين وجدوا في الدراسات اللسانيّة أرضًا خصبة لاختبار آرائهم النظريّة؛ فتفاعلت الآراء والنظريات، للخروج بنتائج تخدم كلا العلمين علم اللسان، والعلم الآخر من هذه العلوم التي شاركته في التنظير، والبحث، والاستنتاج.

وقد فتح النجاح في هذه الدراسات الأبوابَ للانطلاق إلى ميادين أرحب، فدخلت العلوم الرياضيّة، والطبيّة، والحياتية (البايولوجيّة)، والحاسوبيّة، وغيرها من العلوم في حوار مع اللسانيّات، ووجد علماؤها الطريق سالكًا لتجريب نظريّاتهم، بوساطة اللسانيّات؛ذلك أنَّ لغة الإنسان تعبّر عن مخرجات التفاعلات المختلفة في جسم الإنسان، فظهر حقل جديد سُمّي (اللسانيّات التطبيقيّة) أو(اللسانيّات البينيّة)، والحاسوب عقل إلكتروني (اصطناعي) اختُرع لحلّ كثير من المشكلات، له قدرة خارقة على خزن المعلومات، وتبويبها، ومقابلتها، وموازنتها لذا صار الحاسوب مصدر نظريات، ونُظم ومبتكرات، اختصّ بها علماء من المهندسين، والرياضيين، مهمّتهم برمجة الأفكار، والمعلومات، مهما كانت هائلة الحجم، والعدد، وإخضاعها لنظم منطقيّة، ورياضيّة تؤدي إلى نتائج مذهلة تفوق قدرة العقل البشريّ.

وقد تنبّه اللسانيّون في أمريكا في العام 1954 إلى أهميّة توظيف الحاسوب في الدراسات اللسانيّة، وقد تابعهم علماء اللسانيّات في أوربا، في السير في الاتجاه نفسه، ففي عام 1961 فتحوا المراكز المختصّة بذلك، وأجروا الدراسات المشتركة، بفكر لسانيّ وبرمجة المختصين بالحاسوب، وذلك بوساطة تحليل النصوص، وموازنة المتشابهات، والمختلفات و

التحليل الصوتيّ، والصرفيّ، والدلاليّ، والنحويّ، للنصوص، فكانت النتائج باهرة ولاسيّما في مجالي الأصوات والمعجمات.

وقد نقل الدارسون العرب في أوربا هذه الأفكار إلى الوطن العربيّ فكانت مبادرة الدكتور إبراهيم أنيس، بالتعاون مع الدكتور علي حلمي موسى أستاذ الفيزياء في جامعة الكويت في عام 1971 الإنطلاقة العمليّة الأولى للبدء بتوظيف العقل الإصطناعي في الدراسات اللسانيّة، إذ جرى إحصاء الحروف الأصليّة لجذور اللغة العربيّة المستعملة في معجم الصحاح للجوهريّ، تلتها محاولات تطبيقيّة في معجمات عربيّة أخرى، أجرت تحليلات صرفيّة، وصوتيّة، ودلاليّة لاقت استحسان الباحثين لما حقّقته من نتائج مثمرة، في وقت قصير جدّا يتطلّب جهدًا بشريّا طويل الأمد، مع دقّة في النتائج لايحققها الجهد البشريّ.

وقد كانت هذه الحركة الدؤوبة في الوطن العربي من تعدّد إجراء التطبيقات اللسانيّة، وفتح المخابر اللسانيّة، وتأسيس الكليّات والأقسام المختصّة باللسانيّات الحاسوبيّة، في عدد من الجامعات العربيّة قد أثمرت ظهور دراسات تخصّصت بوصف هذه الحركة، وتقويمها، وإظهار اتجاهاتها، ومسارات التفكير في إنضاجها ، فكُتبت بحوث، ونُشرت مقالات وألّفت كتب منها: كتاب الدكتور «نبيل علي» الموسوم بـ (اللغة العربيّة والحاسوب) الذي صدر في عام 1988 وكتاب الدكتور «ذياب عبد العجيلي» الموسوم بـ (الحاسوب واللغة العربيّة) الذي صدر في عام 1996 وكتاب الدكتور«نهاد الموسى» الموسوم بـ (العربيّة نحو توصيف جديد في ضوء اللسانيّات الحاسوبيّة) الذي صدر في عام 2000م.

وهذا الجهد المبارك الذي نَمَا وتطوّر،مازال محدودًا في الوطن العربي لكون الدرس اللسانيّ في مراكزنا البحثيّة العربيّة، يعاني من آفة خطيرة؛هي التبعيّة للفكر الغربيّ في التنظير والتطبيق، وما زالت بنا حاجة لإحداث ثورة فكريّة لتطويع العلوم الأخرى لصياغة نظريات عربيّة في اللسانيّات، وهذه الثورة لها مسوّغاتها وركائزها، في الوطن العربيّ، ولاسيّما في العراق، ويمكن تلخيص تلك المسوغات بوجود الكمّ الهائل من نظريات الفكر اللغويّ العربيّ الذي يصلح أن يكون ركيزة لبناء أفكار، ونظريّات لسانيّة من خلال انتقاء الأفكار

اللسانيّة التي لم يخض في دراستها علماء اللسانيّات في أوربا وأمريكا، وذلك يقتضي توضيحها للمختصّين بعلوم الحاسوب من مهندسين، ومبرمجين ، لبناء برامج متطورة، ترفد الفكر العالميّ بالآراء، والنظريّات، والمخرجات.

وتتطلّب النهضة اللسانيّة التطبيقية الجديدة في العراق تكاتف كلّ الجهود ، وتضافر القدرات الرسميّة، وغير الرسميّة لبناء حركة لسانيّة تطبيقيّة، تسخّر العلوم كلّها لرفد الدرس اللسانيّ بالآراء الجديدة، والنظرات الناضجة، وقد جهرتُ بهذه الدعوة بصوت عالٍ ،وفي بناية وزارة التعليم العالي والبحث العلميّ، من خلال المشروع الذي قدّمتُه إلى معالي وزير التعليم العالي د. «عبد الرزاق العيسى»، ونتج عن تقديمه إقامة ورشة لسانيّة علميّة، في الشهر الثاني من عام 2017 م حضرها السيد الوكيل العلميّ واثنان من مستشاري السيد الوزير، وأقرّ المشاركون في الورشة عددًا من القرارات المهمّة التي تتعلّق بفتح أقسام متخصّصة باللسانيّات التطبيقيّة، وإيجاد دروس لسانيّة في المرحلتين الرابعة والثالثة من الدراسة الأوليّة، وتطوير الدراسات اللسانيّة في الدراسات العليا وغيرها من القرارات التي رُفعت إلى الوزارة ولم يُتخذ بشأنها أيّ إجراء رسميّ.

ولقد أثبت الباحثون العراقيّون جدارتهم في ارتياد هذا الميدان الحيويّ من اللسانيّات الحاسوبيّة، بمبادرات فرديّة ومن خلال عدد من الرسائل الجامعيّة، وقبل الإشارة إلى عدد من الإنجازات العراقية في داخل العراق، لابدّ من الإشارة إلى ما أنجزه العالم العراقي الدكتور سلمان العاني في جامعة أنديانا، من تطبيقات صوتيّة في المختبرات الصوتيّة أدّت إلى نتائج، أغنت كثيرًا من البحوث الصوتيّة؛ التي بُنيت على نتائج الدكتور سلمان العانيّ.

أمّا في العراق فقد استطاع الدكتور «محسن علي عربي» بوساطة استيعابه لبحور العروض العربيّ؛ من إجراء حوارات مع مختصّين بالحاسوب في نهاية التسعينيّات، أدّت إلى وضع برنامج لمعرفة أيّ بحر من بحور الشعر الحرّ، وتفعيلاته، في أيّ بيت شعريّ، وفي المرحلة الثانية من المشروع أدخل الدكتور محسن تعديلات ليكون البرنامج قادرًا على تقطيع أي بيت عموديّ ومعرفة البحر في اللغة العربيّة وفي لغات أخرى كاللغة الكرديّة .

وتمكّنت الدكتورة «جُمانة خالد المشهدانيّ» تضمين أطروحتها للدكتوراه الموسومة بـ (برنامج القراءة الآليّة في اللغة العربيّة دراسة وتحليل) برنامجها الخاصّ بتقويم نُطق الأصوات العربيّة عن طريق موازنتها بنطق قرّاء القرآن الكريم للأصوات، من خلال الصور الطيفيّة للأصوات؛ وذلك بتصميم برنامج يوضح تغيّرات صفات الصوت بشدة الألوان وخفوتها.

وأنجزت الطالبة «سرى هاتف» برنامجًا بالتعاون مع مبرمجة عراقيّة، لبيان دلالات مكوّنات الجملة الشرطيّة في القرآن الكريم أرفقته برسالتها التي أشرفتُ عليها ونالت عليه شهادات من جهات علميّة متعدّدة، ومنها وزارة التعليم العالي والبحث العلميّ.

وهذا التعاون المشترك بين اللسانيّين العراقيّين، والمختصّين بالبرمجة، قد اتسع مداه في العراق كثيرًا، وفي مدة قصيرة، ونتجت عنه برامج كثيرة، وهو أمر دعونا و ندعو كلّ اللسانيّين إلى تبنّيه في دراساتهم اللسانيّة؛ لأنّ المختصّين بالبرمجة متفقّهون في موضوع اختصاصهم؛ غير أنّهم لا يفقهون أسرار اللغة، ولا يعرفون أصواتها، وبناها الصرفيّة، والتركيبيّة، ولا المخرجات الدلاليّة، وسبل إنتاجها، ولا يعرفون كثيرًا ممّا توصل إليه علماء اللغة من نظريات، ورؤى مستنبطة من التحليل اللغويّ، وهذا ما يدعونا إلى التفكير في الموضوعات اللسانيّة الجديرة بالدراسة والتحليل الألكترونيّ، والسعي إلى تحويلها إلى برامج علميّة تطبيقيّة.

ولقد برزت في العراق ظاهرة مهمّة، وإنْ كانت على نطاق ضيّق؛ لكنّها جديرة بالاهتمام، فقد درس عدد من المهندسين العراقيّين اللغة، اختصاصًا آخر غير الهندسة وأكملوا الماجستير، والدكتوراه، في اختصاص اللغة العربيّة، وهذا ما أهّلهم إلى الجمع بين أمرين:

الاختصاص العلميّ في الهندسة الذي يمارسون من خلاله التخطيطَ ، والرسم الهندسيّ، والتصميم ، والبرمجة الحاسوبيّة.

والاختصاص اللسانيّ الذي يتيح لهم الاغتراف من الفكر اللسانيّ والتراث اللسانيّ الأفكار، والنظريّات اللسانيّة، والمزج بين الاختصاصين، بوضع البرامج التطبيقيّة من غير

حاجة إلى خبرة المُبرمجين، إذ يوظّف المهندس الدكتور معرفته بالعلمين لتقديم ما تفتقر إليه الدراسات اللسانيّة، ومن أولئك الذين مزجوا بين العلمين :

المهندس الدكتور «عماد يونس لافي» الذي بنى أطروحته للدكتوراه على توظيف الحاسوب للمقاربة بين الآراء البلاغيّة في اللغة العربيّة، والأفكار المناظرة في الهندسة المعماريّة. وأثبت ذلك من خلال الرسوم الهندسيّة والاستنتاجات الحاسوبيّة ، وحاول تطوير ذلك من خلال بحوثه في النحو وجعل الميدان التطبيقيّ أساسًا لها في تطبيقاته الحاسوبيّة.

وأصدر الدكتور «عادل عباس النصراويّ» وهو مهندس درس اللغة العربيّة اختصاصًا آخر؛ فأكمل الماجستير، والدكتوراه فيها، ووظّف معرفته في الرياضيّات والهندسة بابتكار برنامج رياضيّ في العروض العربيّ، والمعجم، والصوت شرح أفكاره في كتابه الرائع الموسوم بـ (المنهج الرياضيّ في دراسة اللغة والشعر / دراسة صوتيّة، معجميّة، عروضيّة) وهو كتاب حريّ بالدراسة لما يفتحه من أبواب علميّة في الدراسات اللسانيّة.

ولاشكّ في أنّ المهندس الدكتور «كيان أحمد حازم يحيى» أستاذ اللسانيّات في كلية الآداب / جامعة بغداد ، والمهندس الدكتور «عبد الرحمن الجبوري»، والمهندسة الدكتورة زوجته وآخرون قد نهجوا المنهج العلميّ في الدراسات اللسانيّة الحاسوبيّة لمعرفتي بقدراتهم العلميّة الباهرة.

والأمل يحدوني إلى أن تكون هذه الثلّة من اللسانيّين المهندسين الطاقة المبدعة لابتكار دراسات رائدة في اللسانيّات الحاسوبيّة، ترفد الحركة المتطوّرة التي يشترك فيها اللسانيّون، والمبرمجون في ابتكار البرامج اللسانيّة المستندة إلى النظريّات العربيّة في التراث اللسانيّ العربيّ، وهو خزين تفتقر إليه الجامعات الأوربيّة والأمريكية؛ لذا وضعتُ تصوراتي عن مستقبل اللسانيّات التطبيقيّة في العراق في كتابي (هويتنا العربيّة في اللسانيّات ومناهجها) الذي صدر في هذا العام في عمّان في ضوء ما استقر في ذهني من قناعة بأنّ الوقت قد حان لتطوّر هذا النمط من الدراسات في العراق.

ولاشكّ عندي في أنّ الطريق إلى هذا الميدان من الدراسة قد صار سالكًا لبلوغ أعلى مراتب الإبداع والإبتكار في الدرس اللسانيّ العربيّ؛ ليرفد الفكر اللسانيّ العالميّ بما يجعله في تطوّر وارتقاء مستمرين.

وبحوث الندوة التي أقامتها جمعيّة اللسانيّين العراقيّين بالتعاون مع جامعة ذي قار في مدينة الفنون والآداب والعلوم ، مدينة الإبداع والابتكار الناصريّة؛ تعدّ خطوة مباركة لبدء مرحلة جديدة تخطّط جمعيتنا لاجتيازها بنجاح من خلال إقامة ورشات علميّة تعرض من خلالها عددًا كبيرًا من البرامج المستحدثة في هذا المضمار،كما تخطّط الجمعيّة لتحقيق إنجازات كبيرة تضع المنجز العراقيّ في اللسانيّات التطبيقيّة في المكان الذي يليق بتاريخ العراق ومكانته الحضاريّة ومن الله التوفيق.

أ. د كريم حسين ناصح الخالديّ

رئيس جمعيّة اللسانيّين العراقيّين

في الخامس من صفر 1441 هـ

الموافق الرابع من تشرين الثاني 2019م

كلمة اللجنة التحضيرية

هذه محاولة لسانية، وخطوة أعتز بها، وألتمس زادا معرفيا جما للذي يطلع عليها، إذ قلَت نظيراتها في العراق، وقد أسهم فيها مشكورا السيد رئيس جامعة ذي قار الدكتور «يحيى عبد الرضا»، وعميد كلية التربية للعلوم الإنسانية الدكتور «نعيم كريم الشويلي»، جزاهما الله خيرا، وفي الوقت نفسه أشكر رئيس قسم اللغة العربية الدكتور «رحمن فرهود» على جهده الكبير...

وأُثني على الباحثين العلماء المشتركين معنا، وأبارك انعقاد هذه الندوة اللسانية في مدينتنا الحبيبة، مدينة الفكر والإبداع، وادعو للجميع بالتوفيق، وأرجو على المشاركين في الندوة وغيرهم إلى مواصلة السير على النهج القويم في المشاركة الجادة، والفاعلة في حركة تطوير الفكر اللساني العالمي من خلال تقديم بحوث لسانية جديدة فيها إبداع وابتكار، وعدم اعتماد تكرار ما يرد من الغرب وإلباسه أثوابا عربية أو عراقية، لان الفكر اللساني ينتظر منا المشاركة بأفكار جديدة ومناهج جديدة، ونظريات جديدة، وقد آن الأوان لنكون فاعلين في حركة التطور اللساني.

إيمانا من توجه جمعية اللسانيين العراقيين التي أجيزت بموافقة التعليم العالي، لدينا رغبة كبرى بمغازلة اللسانيات التطبيقية، وبيان مساحات اشتغالها، ومن أهم تلك اللسانيات التطبيقية اللسانيات الحاسوبية، التي تقوم على الممازجة بين الحاسوب واللغة، أي إفادة اللغة من علم الحاسوب في أبحاثها لحل إشكالياتها، عبر تنصيب برمجيات خاصة بذلك يقوم بها الحاسوبيون، واللسانيون.

فجاء عنوان الندوة اللسانيات الحاسوبية في العراق المفهوم والمأمول.

لسببين:

- أن العراق فقير بهذا الجانب وهذا مايؤسف حقا، فجاءت هذه الندوة بمثابة وخزة في جعبة الباحثين العراقيين، وبعض الأقسام العلمية في جامعاتنا العراقية.

- أن بعض الباحثين العراقيين لهم فضل السبق ومنهم الدكتور عبد ذياب العجيلي اذ كتب في عام 1996 كتابا بعنوان (الحاسوب واللغة العربية)، ومنهم الدكتور «رعد النواس» إذ كتب في وقت مبكر بحثه في الماجستير، وبقي البحث اللساني «رهين التكرار»، والقِدم، فالركب يسير في دول العالم ونحن نقف بالضد من مسايرة الحياة، والتطورات.

جاءتنا ستة أبحاث بعد أن أعلنا عن الندوة وأعطينا مدة لا تتجاوز ثلاثة أسابيع، ولله الحمد يعد نجاحا قياسا بهذا الزمن، وقياسا بالبحث اللساني في العراق، أشكر القائمين معي على هذه الندوة، في اللجنة العلمية، والتحضيرية.

أشكر الباحثين عناءهم وتعبهم.

أشكر الحاضرين جميعا.

الأستاذ المساعد الدكتور

خالد حوير الشمس

عضو جمعية اللسانيين العراقيين

2019/10/25

مقدمة الكتاب

الحمد لله رب العالمين، حمدا حقا، والشكر له على ما أنعم، ووفق، وهدى..

بعد التطور الفكري الحاصل في ميادين العلوم كلها لم تكن اللغة العربية بمعزل عن ذلك التطور، بل انفتحت بوتقة أبحاثها اللسانية لتمازج بين التصورات اللسانية العامة التي درستها اللسانيات النظرية مع التصورات العلمية الاخرى، فكان نمطا لسانيا يختلف عن غيره في المنهج، والتوظيف، سمي باللسانيات التطبيقية، وهي متعددة منها اللسانيات التربوية أو التعليمية، واللسانيات الاجتماعية، واللسانيات النفسية، واللسانيات الطبية، وغيرها وصولا إلى اللسانيات الحاسوبية، التي دلت على مزج اللغة بالحاسوب وبيان مدى إفادة أبحاث اللغة من معطيات الحاسوب، ومعالجة اللغة، ونصوصها معالجة آلية في مستوياتها المتعددة، فتوجه البحث إلى المعالجة الصوتية للغة، والى المعالجة الصرفية، وإلى المعالجة المعجمية، وإلى المعالجة النحوية، وإلى التحليل النصي والأسلوبي، وغير ذلك جاريا بمعزل عن السياق الذي اعتادت عليه النظرية اللسانية، مما ولد مفاهيم مبتكرة، وجديدة تختلف عن المفاهيم السابقة.

جاء في هذا الكتاب ستة أبحاث تنتمي إلى مجالات متنوعة، بعثت بها الصدفة لأن تخلق الجنبة التكاملية في الطرح، فكان البحث الأول بعنوان (اللسانيات الحاسوبية المفهوم ومجالات التطبيق) لباحثتين من جامعة (حسيبة بو علي) في الجزائر، «عقيلة زموري»، و«صفاء مجاهد»؛ بينتا فيه مفهوم اللسانيات الحاسوبية، ومجال اشتغالها، ثم المساحة التطبيقية لها.

وجاء البحث الثاني بعنوان (اللسانيات الحاسوبية دراسة في التمثلات البينية)، للدكتور خالد حوير الشمس. كلية الآداب/ ذي قار، سجل فيه أبرز نقاط البينية الناتجة عن تلاقح علمي الحاسوب واللغة، ومنها الآلية، والتعبيرية، والتخزين، والتدوين.

ونهد البحث الثالث الجنبة الصوتية في اللسانيات الحاسوبية بعنوان (المعالجة الالية للمستوى الصوتي في اللسانيات الحاسوبية) للدكتورة بشرى عبد الرزاق العذاري كلية التربية / المستنصرية، مسلطة الضوء على أشهر البرامج الحاسوبية المستعملة في المعالجة الآلية

للأصوات، التي تساعد في البحث الصوتي الحديث، وتنسجم، وتتلاءم مع النتائج الصوتية عند القدماء، مع أيسر الطرائق، وأبسطها.

وجاء البحث الرابع ليدرس المساحة الصرفية بعنوان (سبل تطوير اللغة العربية من خلال اللسانيات الحاسوبية- المحلل الصرفي انموذجا) للدكتورة «بختة تاحي» من جامعة شلف/ الجزائر اذ وضحت كيفية عمل المحلل الصرفي بعد التخزين المعلوماتي له.

وانتمى البحث الخامس إلى المجال المعجمي بعنوان (اللسانيات الحاسوبية وفاعليتها في صناعة المعجمية الالي) للاستاذ «اسلام حب الدين»، د. «زينة قرفة» جامعة محمد البشير الإبراهيمي برج بوعريريج/ الجزائر، سعى الباحثان فيه إلى الكشف عن دور اللسانيات الحاسوبية في صناعة المعجم الآلي، وبينا لنا كيف أن اللسانيات الحاسوبية ستثري الخزينة اللسانية المصطلحية العربية آليا، وكيف تساعدنا في وضع معاجم متطورة نجمع فيه تراثنا اللغوي العربي.

وآخر الأبحاث بعنوان (لسانيات المدونة الحاسوبية بين النظرية والتطبيق) للدكتور «خالد شاكر حسن»، كلية الآداب/ ذي قار، كتبه باللغة الانكليزية، مبينا فيه الجنبة الاحصائية وأثرها في تحديد الأساليب، وتشخيص المقاصد، متخذا من نهج البلاغة مدونة مكتوبة.

وينبغي التنويه بأمرين مهمين في هذا الكتاب، تكرار بعض المعلومات، والعنوانات الفرعية في ثناياه، وهذا ليس قدحا به، فلم يكن نتاج أسلوب باحث منفرد، بل هو نتاج استكتاب، وطبيعة الاستكتاب التكرار، وذلك التكرار بحسب طبيعة منهج البحث لكل بحث فيه، ثم إني لفرح بذلك التكرار كون الباحث العراقي له حاجة ماسة هذه الأيام إلى التعرف على المبادئ الأساسية لعلم اللسانيات الحاسوبية.

والأمر الآخر أن هذا الاستكتاب جمع بين الجنبتين النظرية والتطبيقية في معالجاته، وهذه مصدر رصانة له، ومصدر ثراء معرفي تجعل القارئ محيطا إحاطة تامة بين النظري والتطبيقي.

الذي أرجوه من الرب العلي العظيم أن تعم الفائدة في هذه الخطوة العلمية المباركة، وأن يواكب الباحث العراقي الباحثين في العالم، وتقديم الناجز، والناهض لخدمة عربيتنا الكريمة، ولخدمة حياتنا التي صارت بأمس الحاجة لفهم الخطاب الدائر في العالم، ولايفهم ذلك الخطاب إلا بالانطلاق من التحديث اللساني؛ كون البحث اللساني النسقي بات عاجزا عن الفهم، والتفتيق النصي، وبيان مضمرات الخطاب.

وآخر دعوانا أن الحمدُ لله رب العالمين

الدكتور خالد حوير الشمس

عضو جمعية اللسسانيين

5/ تشرين الثاني / 2019

اللسانيات الحاسوبية المفهوم ومجالات التطبيق

الباحث(ة): عقيلة زموري

الباحث(ة): صفاء مجاهد

جامعة حسيبة بن بوعلي- الشلف- الجزائر

ملخص:

اللسانيات الحاسوبية فرع من اللغويات التّطبيقية يجمع اللغة بعالم الآلـة، تعتمـد فيـه آليات واستراتيجيات ذات أبعاد رقمية رياضية، يتعامل بها مهندسـي الإعـلام الآلـي و أهـل اللغة من أجل تحقيق نوع من الاتـصال اللغـوي الطبيعـي باللغـات الاصـطناعية، شـهد هـذا الميدان العديد من المجالات و البرامج التطبيقية التي عُنيت بالتدقيق والتّحليـل اللغـوي الآلـي كالمدقق الإملائي، والترجمة الآلية، والتحليل النحوي وغيرها، وعليه فإناّ نروم في هذه الورقـة البحثية الحديث عن مفهوم اللسانيات الحاسوبية في الوطن العربي ،وكذا تسليط الـضوء علـى مجلاتها التطبيقية ومدى فاعليتها في حوسبة اللغة.

Summary:

Computational linguistics is a branch of applied linguistics that combines language with the machine world. It adopts mechanisms and strategies of numerical mathematical dimensions, in which computer engineers and people of language deal in order to achieve a kind of natural linguistic communication in synthetic languages. In this paper, we want to talk about the concept of computational linguistics in the Arab world, as well as highlighting its applied journals and their effectiveness in computerization. Language.

key words:

Computational linguistics - Spell checker - Machine translation - Grammar analysis

تمهيد:

شهدت اللغة في العصر الحديث حركية واسعة بسبب ارتباطها بالتكنولوجيا رغبة من علمائها في مسايرة التطور الحضاري وحمايتها من الضياع، الأمر الذي أدى بهم إلى تركيز جهودهم من أجل خلق بعد تكاملي يجمعها بعالم الرّقمنة وفي محاولة لتطويع الأجهزة لحوسبة هذه اللغة ومعالجتها بذكاء يقارب الفكر الإنساني في تحليلة وتعامله . كان ميدان اللسانيات الحاسوبية أحد الفروع التطبيقية بوصفها مُعوّلاً أساسياً في معالجة اللغة. فما المقصود باللسانيات الحاسوبية؟

وما مدى فاعلية تقنياتها العملية في معالجة اللغة؟.

اللسانيات الحاسوبية: Computational linguistics

المتأمّل في هذا التركيب يدرك يقينا أنّه سيجمع دلالة الوحدتين في مفهوم واحد يحدّد قيمته التكاملية، انطلاقا منه عمد الباحثون إلى اطلاق مفاهيم وتعريفات تخص هذا الميدان فوجدوا أنّ:

فكرة هذا المفهوم في تركيبته المزجية التي تتبنى ربط اللسانيات بالتقنية الحديثة المتطورة عبر ما بات يعرف بالهندسة المعلوماتية، التي تستعير من المنطق الرياضي النموذج الفعّال في معالجة اللغات الطبيعية سواء كانت غربية أم عربية، وبطبيعة الحال هذه الأخيرة ما زالت تسعى بكل ما تملكه لمواكبة العصر التكنولوجي المتقدم.

اللسانيات الحاسوبية Computational linguistics هو نظام بيني بين اللسانيات نظام بيني، بين اللسانيات وعلم الحاسوب المعني بجوسبة الملكة اللغوية، وهي تنتسب إلى العلوم المعرفية وتتداخل وحقل الذكاء الاصطناعي وهو فرع من علم الحاسوب يهدف إلى وضع نماذج حاسوبية للإدراك الإنساني وللسانيات الحاسوبية مكونان :

تطبيقي ونظري، أما التطبيقي فأولى عنايته بالناتج العملي لنمذجة الاستعمال الإنساني للغة ، وهو يهدف إلى إنتاج برامج ذات معرفة باللغة الإنسانية، تستند الحاجة إليه أجل تحسين التفاعل بين الإنسان والآلة؛ إذ إن العقبة الأساسية في طريق هذا التفاعل بين الإنسان والحاسوب إنما هي عقبة التواصل ، وحواسيب هذه الأيام لا تفهم لغتنا، أما لغات

الحاسوب فيصعب تعلمها كما أنها لا تطابق بنية التفكير الإنساني[1]، أو هي" فرع من فروع اللّسانيّات الحديثة. غايته اللغة ووسيلته الحاسوب، وتمثّل ذلك العلم الذي يُعْنَى بإخضاع الآلة لمتطلبات البحث اللغوي على كافة مستوياته استنادا إلى مُعطيات اللغات الطبيعيّة، ويغلب عليه الجانب التنظيري، أما معالجة اللغات الطبيعيّة فتُعنى بتوجيه الأنظمة الحاسوبيّة إلى محاكاة قواعد اللغات الطبيعيّة اعتمادا على معطيات هذه اللغات، وتمثل الجانب التطبيقي من اللّسانيات الحاسوبيّة. وتشير تقنيّات اللغات الإنسانية إلى مجموعة التطبيقات التي تستخدم في بناء وتطوير أدوات المعالجة الآلية للغات الطبيعيّة"[2].

يرى الدرس اللساني الحديث أنّ مصطلح اللسانيات الحاسوبية أعطاه عدد من الباحثين مصطلحات أخرى مثل الإعلامية، واللسانيات المعلوماتية، وعلم اللغة الحاسوبي، ويعود الاختلاف إلى ضبابية تمثل علماء اللغة العرب لصلة علوم الحاسب بما هي علوم تقنية بالعلوم اللغوية بما هي علوم إنسانية، فهي في الأخير علوم أساسها نشاط من أنشطة الإنسان[3].

ومعنى هذا أنّ الحاسوب عبارة عن أداة تُعين الباحث في اللغة على استثمار أفضل بمحاولة الكشف عن الأشياء الدقيقة بقدرة تفوق العقل البشري؛ وكمثال على ذلك القيام بعمليات إحصائية غاية في الدقة فيما يخص العمل المعملي،وطريقة الترتيب التي يبرمج بواسطتها بحيث يكون الحاسوب أسرع بكثير من الإنسان...في استنتاجاته وملاحظاته؛ ولعلّ الحديث عن هذه العلاقة بين اللغة والحاسوب أشبه بعلاقة الوسيلة بالأداة[4].

---

(1) نهاد الموسى، العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، المؤسسة العربية للدراسات والنشر، الأردن، ط1، 2000، ص53.

(2) المعتز بالله السعيد، علم اللغة والتقنيات المعاصرة، 15 أكتوبر 2014، ص: 07، moataz@cu.edu.eg

(3) ينظر صالح بن هادي رمضان، التفكير البيني أسسه النظرية وأثره في دراسة اللغة العربية وآدابها، ص155.

(4) ينظر:أحمد زد محبك ابن مصطفى :الحاسوب وتنمية المقدرة اللغوية عند الطفل،مجلة المجمع الجزائري للغة العربية،العدد 7،جوان 2008 م،ص43

مجالات اشتغال اللسانيات الحاسوبية:

تشمل اللسانيات الحاسوبية عدة مجالات تمثل حقل اشتغالها، وقد خلف تطبيقها في ذلك منجزات قيِّمة، مِن أهمها ما يلي:

تعدّ تقانات المعالجة الآلية للغات هي البناء التحتي الأساسي الواجب توفيره حتى يتم فوقه بناء كافة التطبيقات الأخرى، ولا تقف تقانات المعالجة الآلية حاليا عند حد إدخال البيانات والمعلومات واسترجاعها، بل تمتد إلى التطبيقات اللغوية التخصصية من تدقيق إملائي، وتدقيق نحوي، وتحليل صرفي، وترجمة آلية، و تقانات التعرف الضوئي على الكتابات الطباعية، واليدوية، والتعرف على الكلام وقراءة النصوص، الإحصاء اللغوي، الحاسوب والمعجم... [1].

1- المدقق الإملائي:

يهدف إلى تعرف الحاسب على بنية الكلمة العربية من خلال القواعد التي يضعها اللسانين لا من خلال معجم الكلمات المخزنة، الشيء الذي يظهر أهمية الخبرة اللسانية في الصناعة اللغوية.

2- المدقق النحوي:

يدقق في النص المكتوب؛ من أجل الوقوف عند الأخطاء الهجائية والنحوية فيه، لكي لا يتم الحكم على عدة جمل وكلمات صحيحة لغويا بالخطأ نتيجة ضعف العتاد اللساني...

---

(1) لطوف عبد الله، عبد اللطيف عبيد، استخدام التقانات الحديثة في تطوير اللغة العربية ، المنظمة العربية للتربية والثقافة والعلوم إدارة التربية، تونس ، دط،2010، ص29. وينظر محمد محمد داود :العربية وعلم اللغة الحديث، دار غريب للطباعة والنشر،القاهرة، مصر، ص273، وينظر منصور الغامدي، مدخل إلى اللسانيات الحاسوبية، الرياض، المملكة العربية السعودية،ط1، 2018، ص7.

3- المشكّل الآلي:

يقوم بضبط النص العربي على مستوى الشكل بناءً على ما خُزّن من قواعد عربية في ذاكرة الحاسب وبالتالي، فمن خلال هذه البرامج اللسانية التي يمكن أن ندرجها ضمن المجال التركيبي، يتضح جليا ما تتطلبه الهندسة اللسانية من معرفة لسانية أو لا، ثم معرفة حاسوبية ثانيا، وهي ثنائية لا تقبل الفصل ولا المجاورة[1].

4- التحليل الصوتي:

"تكاد اللغة في مظهرها أن تكون أصواتا إنسانية، يحللها عالم الأصوات اللغوية ويصفها،كما يشرح لنا كيفية صدورها، وأعضاء النطق التي تساهم في إخراجها، وقد استطاع المحدثون بعد تجارب كثيرة ودراسات مستفيضة ورحلات طويلة أن يجمعوا لنا الكثرة الغالبة من تلك الأصوات الإنسانية، وأن يصفوها وصفا دقيقا، ويسجلوا منها نماذج منطوقة فوق أشرطة وأسطوانات ثم رموز لكل منها برمز خاص اصطلحوا عليه، وقام لديهم بمثابة رسم عالمي وهكذا نظروا إليه نظرة عالمية، بصرف النظر عما ينتمي إليه الصوت من اللغات"[2].

"فالمستوى الصوتي تم معالجته آلياً بوساطة تحليل طيف الصوت، وتوليد (إنتاج) الكلام وتخزين الأنماط الصوتية للشخص المتكلم .وتبعاً لهذا تمَّ تصميم أجهزة تخليق الكلام وتحليله، وتوليد الكلام المنطوق آلياً بتحويل النصوص المدخلة في جهاز الحاسوب إلى مقابلها الصوتي، وعلاج عيوب النطق"[3]. "ولقد أفادت الدراسة المخبرية في تعيين الصفات الفيزيائية للصّوت اللغوي داخل البنية اللّغوية بما يحدث فيها من تأثير وتأثّر بين الأصوات، وذلك

---

(1) ابراهيم مهديوي، اللسانيات الحاسوبية رقمنة اللغة العربية ورهان مجتمع المعرفة، شبكة الألوكة ، 16/ 11/ 2016، www.alukah.net/literature

(2) ابراهبم أنيس، من أسرار اللغة، مكتبة الأنجلو المصرية، القاهرة، ط3، 1966، ص4.

(3) عبد الرحمان حسن بن عارف ، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية جهود ونتائج،جامعة أم القرى ص15.

لوضع أنموذج صوتي يتكون من توصيف دقيق للصوت المرجعي (الثابت) واختلاف طرق أدائه من ناطق إلى آخر عن طريق المحلل الصوتي"[1].

5- التحليل الصرفي:

يعد علم الصرف أساسا متينا في التعامل مع اللغة العربيةُ، أما حوسبته فتعد مطلبا أساسيا لكثير من التطبيقات اللغوية التي تعتمد على النظم الآلية، إذ يبحث علم الصرف في التغييرات التي تطرأ على بناء الكلمات وصورها المختلفة من الداخل وأسس ترتيب عناصرها وظواهر التفاعل بين هذه العناصر،كما يتيح معرفة مكان الحرف الأصلي، ومكان الحرف الزائد في الكلمة، بالإضافة إلى الاهتداء إلى أصلها الاشتقاقي والمعجمي.

إن التحليل الصرفي الآلي يقصد به العملية العكسية للتوليد أو التوليد الصرفي الآلي حيث تقوم العملية الحاسوبية بتحليل الكلمات إلى عناصرها البنائية أو الصرفية وتحديد مختلف صفاتها آليا، وتشتمل على صفات البناء الصرفي والخصائص النحوية والدلالية والصوتية للكلمات، إذ يتم بواسطتها تحديد العناصر المكونة للكلمة من سوابق ولواحق وجذع الكلمة وجذرها وأنماطها الصرفية وعمليات الإبدال والإعلال وغير ذلك، وتعد عمليات التحليل والتوليد الصرفي أساس بناء مختلف التطبيقات اللغوية الحاسوبية.

يقصد بميكنة التحليل الصرفي قيام النظام الآلي باستخلاص العناصر الأولية لبنية الكلمة، وتحديد سماتها الصرفية والصرف نحوية والدلالية والتي يمكن استنباطها من بنية الكلمة، ومن تمثيل ذالك فيما يلي:

تحليل كلمة "وبإيصاله"

السوابق: "و"حرف عطف، "ب" حرف جر

جذع الكلمة "إيصال"

الرتبة النحوية: مصدر ثلاثي مطرد

---

[1] راضية بن عريبة، الصوت اللغوي والحوسبة الآلية، جامعة حسيبة بن بوعلي كلية الآداب واللغات، قسم اللغة العربية، الأكاديمية للدراسات الاجتماعية والإنسانية،ص1. http://www.webreview.dz/spip.php?article1460

قسم الكلم: اسم جماد، محسوس، قابل للعد، أو اسم مجرد غير قابل للعد

جذر الكلمة: وصل

الصيغة الصرفية: إفعال

الميزان الصرفي: إيعال

الحالة التصريفية: مفرد ، مذكر، مجرور

العلامة الإعرابية: الكسرة الظاهرة

اللواحق: ضمير الملكية للمفرد الغائب (ه)

عمليات التعديل الفنولوجي: إبدال فاء الجذر (الواو في وصل، ياءه، ومماثلة حركة الضمير المتصل (هاء الغائب) مع علامة الإعراب التي تسبقه الكسرة[1].

ومما لاشك فيه أن معالجة الصرف العربي آليا يعد من أهم وأوجه الحقول الواجب العناية بها لبناء نظم لغوية حاسوبية عربية، حيث يمثل الصرف اللبنة الأساس التي تنبني عليها مختلف النظم اللغوية الأخرى ، ومن الواضح أن أهمية المعالجة الآلية للصرف العربي تكمن في التعرف على الاستخدامات الممكنة التي من أهمها: ضغط النصوص وتشكيلها وتشفيرها، واسترجاع النصوص وتحليلها، وتمييز الكلام وتوليده، وتصحيح الأخطاء الإملائية، والترجمة الآلية، والتعليم بالحاسوب.

6- التحليل الدلالي:

يقوم المحلل الدلالي بتفسير الجملة من الناحية الدلالية، ومدى قبولها أو عدم قبولها دلاليا ، وذلك بعد القيام بعدة خطوات تتضمن عمليات استدلال واستنتاج بالاعتماد على بنى الحقول الدلالية، وذلك لاختيار الإعراب الصحيح لها انطلاقا من دراسة القواعد الكلية ومعرفة أنواع الدلالات للكلمة وأنواع المعاني[2].

---

(1) نبيل علي: اللغة العربية والحاسوب، التعريب ، 1988، ص:306.

(2) سمر معطي، فاضل سكر، معالجة اللغة العربية آليا باستخدام الذكاء الصنعي،وقائع الندوة الدولية الثالثة حول المعالجة الآلية للغة العربية،4-5 ماي 2009، المدرسة المحمدية للمهندسين،الرباط، المغرب،ص17.

يمثّل توصيف النظام الدلالي للغة العربية نواة المعالجة الآلية ذلك أنّ أغلب المعالجات التي تختص بها التراكيب اللغوية تعتمد بشكل كبير عليها، ويعد من أعسر المباحث تناولا نتيجة تعدد المعاني الدلالية التي يتعسر على الحاسوب فهمها بالمقارنة مع الإستيعاب الذهني للإنسان حيث يستطيع بكل سهولة التعرف على المعنى بإختلافه سواء كان (معجمي، صرفي، نحوي ...) عكس العقل الإلكتروني(1).

يهدف البحث الدلالي إلى الكشف عن المعاني في اللغات الإنسانية وإلى معرفة العلاقات التي تربط بين أجزاء المعنى الواحد وما ينطوي تحته من مضامين العلاقات التي تصل إلى معنى بآخر، وعلى الرغم من اختلاف الثقافات البشرية والتجمعات الإنسانية، واختلاف الزمان والمكان فإنّ الطريقة التي يتكون منها المعنى واحدة(2).

فالمعنى من هذا المفهوم هو الهدف الأسمى للبحث الدلالي لتحديد والكشف عن المعاني للغات الإنسانية رغم اختلاف الثقافات والسياقات التي ترد فيها اللفظة الواحدة.

7- التوثيق:

يُعد أحد أهم مجالات تطبيق اللسانيات الحاسوبية، ذلك أنه يتمثل في عملية جمع المعلومات وتسجيلها وتنظيمها وفق أساليب حديثة، أو هو كما يعرفه «برجس عزام»: "شكل من أشكال العمل الببليوغرافي الذي يستخدم وسائل متعددة؛ كالكشافات والمستخلصات والمقالات الببليوغرافية، إضافة إلى الوسائل والطرق التقليدية الأخرى؛ كالتصنيف والفهرسة؛ وذلك لجَعْل المعلومات سهلةَ المنال، والوصول إليها سهلًا أيضًا"(3). لكن بفضل الثورة التكنولوجية صار الحاسوب الوسيلة الرئيسة والفعّالة التي تستعمل في التوثيق؛ حيث سهل كثيرا عمليات تخزين المعلومات وأرشفتها كيفما كان نوعها، ومهما بلغ حجمها، كالتصنيف الأتوماتيكي للملفات والمؤلفات تبعًا لفهارسها أو مؤلفيها أو مواضيعها وبالتالي

(1) ينظر وجدان محمد صالح كنالي: اللسانيات الحاسوبية العربية الإطار والمنهج ،المعهد الإسلامي للبحوث والتدريب ،ص16 .

(2) ينظر سمير أستيتية ، اللسانيات المجال والوظيفة والمنهج ،عالم الكتب الحديث للنشر والتوزيع ،ص 257.

(3) برجس عزام، مراكز المعلومات، دار جليل، دمشق، ط1، 1992 ، ص25.

سهل استغلالها و تبادلها بين المستخدمين في مختلف المجالات وبين الشعوب والأمم بصفة عامة.

8- صناعة المعاجم الإلكترونية والبنوك المصطلحية:

تعمل اللسانيات الحاسوبية على تطوير العمل المصطلحي و بناء المعاجم الإلكترونية من أجل تسهيل وتخزين المادة العلمية وتعميم استعمالها ونشرها وتداولها، ويتطلب ذلك إلمامًا كبيرًا بجميع مستويات اللغة، فيتم الاعتماد على أدوات مُشفَّرة، وقاعدة معطيات مُرمَّزة تخص مستويات اللغة صرفا ونحوًا ودلالة؛ بحيث تخضع المادة المعجمية في المعجم الإلكتروني لبناء يلم إلمامًا دقيقًا بفروع اللغة، كما يشترط فيه أن يكون شاملًا وعامًّا؛ ذلك أنَّ البرنامج اللساني المُعَدَّ للمعالجة الآلية لا ينبغي أن يفشل في العثور على أية معلومة كيفما كان نوعها وكيفما اتفق؛ لأن أي خطأ في المعلومات المدخَلة من شأنه أن يتسرب إلى باقي مفردات الجملة، أو كل النص برمته، ومن ثمة يعرقل عملية اشتغال البرنامج"[1]، كما يجب أن تكون المعلومات في المعجم الإلكتروني واضحة موائمة للمداخل المعجمية المراد معالجتها آليًّا.

أما بالنسبة للبنوك المصطلحية، فإن اللسانيات الحاسوبية تعمل على إنشائها انطلاقًا من تخزين المصطلحات مُرفَقة بمعلومات عن كل مصطلح مفرد، فتسمح هذه البنوك المصطلحية بتخزين معطيات دقيقة عن كل مصطلح في ضوء نصوص موثقة، مع ذكر ما يقابله بلغات متعددة، وتوضيح مجالات استخدامه، وأساليب توظيفه، وكذا الإشارة إلى مرجعه، سواء كان معجمًا أو معهدًا علميًّا، أو نصًّا، أو وثيقة، فتمكن بذلك وتطور الترجمة الآلية وبناء المعاجم المختصة مما يمثل أداة طيعة ووسيلة خادمة للمتعلمين والمحررين والمترجمين بصفة فعالة.

ومن هنا تتضح العلاقة بين المعجم الإلكتروني والبنوك المصطلحية، وذلك بوساطة تنظيمه لها، بصفة عجيبة وطريقة لا يتأتى لذاكرة الإنسان المحدودة القيام بها.

---

[1] عمر مهديوي ، توليد الأسماء من الجذور الثلاثية الصحيحة في اللغة العربية - مقاربة لسانية حاسوبية-، ص 21.

9- الترجمة الآلية:

تعد الترجمة الآلية أحد أهم مجالات اللسانيات الحاسوبية على الإطلاق وتقتضي الترجمة الآلية نقل النصوص من اللغات الأصلية المصدر إلى اللغات الفرعية الهدف مع مراعاة خصوصيات النقل التي تفرضها اللغة المستهدفة.

وتعد اللغة الإنجليزية، كما سبق الإشارة إليها، اللغة الطبيعية الأولى التي خضعت لهذه العملية، أي إلى أولى تجارب الترجمة الآلية.

وأما عمل الحاسوب في عملية الترجمة الآلية باعتباره أهم ركيزة فيها فإنه يُزوِّد المستخدم بالترجمة المطلوبة إن كانت مسجلة في ذاكرته، وفي حال غيابها يقترح عليه مقابلًا لذلك؛ وهذا ما يفرض توفير المصطلحات والتراكيب المتقاربة وتسجيلها في ذاكرة الآلة، نظرًا لتعدُّد معاني ومقاصد الألفاظ في اللغات الطبيعية، ولارتباطها بالسياق الاستعمالي التداولي والتركيبي، وعليه فإن نجاحَ الترجمة الآلية الدقيقة والمطلوبة رهين التدخل البشري المطلوب لتوضيب النص قبل ترجمته، أو تهذيبه بعد ترجمته، أي أنه يستوجب تظافر خبرة اللسانيِّين العارفين باللغة أكثر من غيرهم في قواعدها؛ نحوًا ودلالة وتركيبًا وصرفًا، وخبرة الحاسوبيين المهتمين بالمجال المعلوماتي الثقاني في تصميم برامج وأنظمة لتوصيف اللغات الطبيعية بشكل يتماشى ومتطلبات عصر التكنولوجيا والاتصالات، من أجل مواكبة الحضارات العالمية المتقدمة في المجال اللغوي الآلي.

ويبقى للترجمة الآلية دورها الفعال في توظيف المعرفة العلمية والتقنية في المجتمع لسد الفجوة العلمية الناتجة عن تضخم الإنتاج العالمي واتصال الثقافات في عصر العولمة.

10- إنتاج النصوص:

لما كان الحاسوب محاكيا لذهن البشري، فإن اللسانيات الحاسوبية تجعل منه خيرَ بديل للإنسان في إنتاج النصوص، بحيث يعتمد في ذلك على برامج وأنظمة مثبّتة فيه تُمكنه من إعادة تصحيح ومراجعة النصوص وتخزينها في ملفات مغلقة تُفتَح كلما دعت الحاجة إلى

ذلك، ومن ثمَّ، فالحاسوب وفر إمكانية الإنتاج النصي المتعدد مثل المراسلات، والدعوات، والتعليقات الصحفية، والإعلانات الإشهارية، وما شاكل ذلك.

11- تعليم اللغات والتعريف بالثقافات:

لعل أبرز مجال للسانيات الحاسوبية بوصفها علما تطبيقيا، مجال تعليم وتعلم اللغات والثقافات، حيث اعتمدت لتلك الغاية على استعمال الحاسوب باعتباره من أهم الوسائط التي تمزج بين الصوت والصورة والكتابة، والهدف الأسمى من هذه العملية هو تجاوز الطرق التعليمية التقليدية القائمة على "التلقين والتحفيظ والتسميع؛ أسلوبًا أساسيًّا ورئيسيًّا في نقل المعرفة"[1]، وإيجاد طرق جديدة تمكن من استغلال قدرات الحاسب من لدن المستخدمين، الشيء الذي أدى إلى إعداد برامج حاسوبية تعليمية، تتماشى مع النظريات البيداغوجية والتعليمية الراهنة.

وقد عرف هذا المجال إنجاز أعداد هائلة من البرامج التعليمية بصفة عامة والمتعلقة بتعليم اللغات بصفة خاصة، كما حظي المجال الثقافي بنفس الأهمية من خلال إعداد مجموعة من البرامج والأنظمة التثقيفية، التي تسعى إلى تنمية الرصيد المعرفي الثقافي للمتعلمين.

---

(1) السيد، محمود أحمد، اللسانيات وتعليم اللغة، دار المعارف للطباعة والنشر، سلسلة الدراسات والبحوث المعمقة،ط1، العدد9، سوسة – تونس، ص، 137.

خاتمة:

خلص البحث إلى مجموعة من النتائج أهمها:

1- مدى مساهمة المحددات اللغوية و تطبيقاتها في معالجة اللغة.

2- ضرورة التدقيق في توصيف اللغة من أجل بناء نماذج وبرامج آلية تُعنى بميكنة اللغة من خلال تمكنها من كل القضايا الدقيقة في اللغة.

3- ضرورة تكاثف الخبرات اللسانية والبرمجية من أجل توسيع مجالات اللسانية الآلية،وتسهيل عملية توطين المعرفة اللغوية ، وكذا العمل على إنشاء برامج قابلة للتحيين بغية استدراك بعض القواعد التي يمكن أن تكون قد أهملت أو لم تعالج بدقة.

قائمة المصادر والمراجع:

1- عبد الرحمان حسن بن عارف، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية جهود ونتائج، جامعة أم القرى.

2- منصور الغامدي، مدخل إلى اللسانيات الحاسوبية، الرياض، المملكة العربية السعودية، ط1، 2018.

3- إبراهيم أنيس، من أسرار اللغة، مكتبة الأنجلو المصرية، القاهرة، ط3، 1966.

4- إبراهيم مهديوي، اللسانيات الحاسوبية رقمنة اللغة العربية ورهان مجتمع المعرفة، شبكة الألوكة، 16/ 11/ 2016، www.alukah.net/literature

5- أحمد زد محبك ابن مصطفى: الحاسوب وتنمية المقدرة اللغوية عند الطفل،مجلة المجمع الجزائري للغة العربية، العدد 7، جوان 2008م.

6- السيد، محمود أحمد، اللسانيات وتعليم اللغة، دار المعارف للطباعة والنشر، سلسلة الدراسات والبحوث المعمقة، ط1، العدد9، سوسة تونس.

7- المعتز بالله السعيد، علم اللغة والتقنيات المعاصرة، 15 أكتوبر 2014،moataz@cu.edu.eg

8- برجس عزام، مراكز المعلومات، دار جليل، دمشق، ط1، 1992.

9- راضية بن عريبة، الصوت اللغوي والحوسبة الآلية، جامعة حسيبة بن بوعلي كلية الآداب واللغات، قسم اللغة العربية، الأكاديمية للدراسات الاجتماعية والإنسانية، http://www.webreview.dz/spip.php?article1460.

10- سمر معطي، فاضل سكر، معالجة اللغة العربية آليا باستخدام الذكاء الصنعي، وقائع الندوة الدولية الثالثة حول المعالجة الآلية للغة العربية، 4-5 ماي 2009، المدرسة المحمدية للمهندسين، الرباط، المغرب.

11- سمير أستيتية، اللسانيات المجال والوظيفة والمنهج، عالم الكتب الحديث للنشر والتوزيع.

12- صالح بن هادي رمضان، التفكير البيني أسسه النظرية وأثـره في دراسـة اللغـة العربيـة وآدابها،الرياض، المملكة السعودية، د.ط، 1432هـ.
13- عمر مهديوي، توليد الأسماء من الجذور الثلاثية الصحيحة في اللغة العربيـة – مقاربـة لسانية حاسوبية–.
14- لطوف عبد الله، عبد اللطيف عبيد، استخدام التقانات الحديثة في تطوير اللغـة العربيـة ، المنظمة العربية للتربية والثقافة والعلوم إدارة التربية، تونس، دط، 2010.
15- محمد محمد داود: العربية وعلم اللغة الحديث، دار غريب للطباعـة والنـشر، القـاهرة، مصر.
16- نبيل علي: اللغة العربية والحاسوب، التعريب، 1988.
17- نهاد الموسى، العربية نحـو توصـيف جديـد في ضـوء اللـسانيات الحاسـوبية، المؤسـسة العربية للدراسات والنشر، الأردن، ط1، 2000.
18- وجدان محمد صـالح كنـالي: اللـسانيات الحاسـوبية العربيـة الإطـار والمـنهج، المعهـد الإسلامي للبحوث والتدريب.

اللسانيات الحاسوبية دراسة في المعطى البيني

الاستاذ المساعد الدكتور خالد حوير الشمس

جامعة ذي قار / كلية الاداب

ديباجة البحث

نجد الحضور الحاسوبي في العلوم كلها، في ظل رهانات الحداثة، فلا استغناء عنه في علم الاقتصاد، وعلم الإدارة، أو علم الكهرباء، وغير ذلك، فهو في كل مكان وفي العلوم كلها، بضمنها اللغة؛ إذ فرض نفسه في كل مفصل منها؛ لتواكب التطور المعرفي، والتكنولوجي، والبحث عن وسائل السرعة، والراحة.

هذا التشكيل الالكتروني يدعونا أن نطمح إلى دراسة البينية الآلية في اللغة، للوقوف على بعض المصطلحات، والمفاهيم، وبيان معناها في العلمين، ومدى التلاقح بينهما، والعلاقة، وإظهار المفردات، والمتناولات، وما يتعلق بعلم الحاسوب ووجوده في دراسة اللغة بلحاظين، الأول: بوصفها منطوقا أو ملفوظا، واختلافها بطريقة التوظيف، بوصفها استعمالا آليا أو ميدانيا، ومن ثمَّ اللحاظ الآخر: التوظيف التكلمي، فقد نستعمل الحاسوب عبر النطق، والكتابة، حينما ننطق، فيتحول المنطوق إلى مكتوب، أو نكتب، فتتحول الكتابة إلى صوت عبر البرامج المعدة لذلك.

لأوضح أن البينية قارة في اللسانيات من هذه الجهة الحاسوبية، ولتكون طريقة التناول مميزة عن غيرها من الدراسات، فقد وجِد في المكتبة العربية دراسة اللسانيات الحاسوبية من جهة النشأة وما يتعلق بها، مثل دراسة نبيل علي، التي فرشت مساحة واسعة عن مسار هذا الفرع اللساني، ودراسة نهاد الموسى، التي وصّفَت العربية حاسوبيا، ودراسات عبد الرحمن الحاج صالح، التي أظهرت محاولات الخليل الحاسوبية.

متن البحث:

لاشك أن الأساس الفلسفي، الذي انطلقت منه اللسانيات الحاسوبية هي ثورة المعلومات، والتكنولوجيا، فقامت على أساس ثنائية الإدخال والإخراج لما يوصَّف توصيفا دقيقا، وهنا تحتاج إلى كفايتين لغوية وحاسوبية، وتقوم الكفاية اللغوية على التمظهر اللساني، ووسم هذا التمظهر اللساني بالمكون النظري، كونه يتناول النظريات الصورية للمعرفة اللغوية، التي يحتاجها الإنسان في فهم اللغة، بينما تقوم الكفاية الحاسوبية على البناء التقني الذي يؤدي ما يؤديه الإنسان في توصيف الظواهر اللغوية[1]، ووسم بالمكون التطبيقي، كونه يهدف إلى إنتاج برامج ذات معرفة باللغة[2].

فيعاني الدارس صعوبات؛ بسبب ابتعاد الحقلين كثيرا، وافتقار اللسانيين إلى اختصاص الحاسوب، وابتعاد اللغة التي ندرسها في ضوء اللسانيات العامة، وبعض اللسانيات التطبيقية عن مقولات هذا الحقل الالكتروني، فصار علما حديثا، يحتاج إلى تأن حتى يتضح، وينضج، فبقيت المكتبة العربية في شحة منه مقارنة باللغة الإنكليزية، التي سبقت لغتنا الشريفة في دراسة الأثر الحاسوبي أو المعلوماتي على اللغة.

تعددت مفاهيم اللسانيات الحاسوبية باستراتيجيات عدة، منها انطلاقا من الأس البيني، والجنبة الرياضية فيه، فقيل: ((وظاهر ظهورا جليا أن هذا العلم فرع بيني، ينتسب نصفه إلى اللسانيات وموضوعها اللغة، ونصفه الآخر حاسوبي، موضوعه ترجمة اللغة إلى رموز رياضية، يفهمها الحاسوب، أو تهيئة اللغة الطبيعية لتكون لغة تخاطب، وتحاور مع الحاسوب، بما يفضي إلى أن يؤدي الحاسوب كثيرا من الأنشطة اللغوية التي يؤديها الإنسان، مع إقامة الفرق في الوقت والكلفة)) [3].

---

(1) ينظر: توظيف اللسانيات الحاسوبية في إعادة توصيف قواعد النحو العربي، إبراهيم إسماعيل المصري، بحث منشور في مجلة كيرالا، العدد 11، جامعة كيرالا، الهند، 2018، ص 266.

(2) ينظر: العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، نهاد الموسى، 53- 54.

(3) اللسانيات الحاسوبية العربية المفهوم التطبيقات الجدوى، وليد أحمد العناتي، مجلة الزرقاء للبحوث والدراسات، المجلد 7، العدد الثاني 2005، ص 62.

وقد جمع نوفس الأساسين في تعريفه، الأساس الرياضي واللغوي، فقال: ((بأنها فرع عن علمي اللغة والحاسب، يهدف إلى تصميم نماذج رياضية للتراكيب اللغوية؛ لتمكن من معالجة اللغة آليا عن طريق الحاسب... على أنه تشكيل للنظريات، والنماذج اللغوية، وتنفيذها على الآلة، ويرى أن بإمكاننا النظر إليه على أنه وسيلة لتطوير نظريات لغوية جديدة بمساعدة الحاسب)) [(1)].

وقد وضع «ديفيد كريستال» تعريفا نحى فيه منحى التفصيل عن بعض مجرياتها، فقال: ((هي فرع من الدراسات اللغوية التي توظف فيها التقنيات والمفاهيم الحسابية، وتوضيح المشكلات اللغوية والصوتية، وتطوير المجالات البحثية، بما فيها إنتاج أصوات كلامية بوسائل اصطناعية عن طريق توليد الموجات الصوتية، وتردداتها، وتمييز الكلام، والترجمة الآلية، ورقمنة الأبجديات، وإجراء اختيارات قواعدية، إضافة إلى مجالات أخرى تستدعي الإحصاء، والتحليل)) [(2)].

وقد مالت بعض التعريفات نحو الابتسار، والاختصار، ((وهو علم حديث يستخدم الحاسوب في دراسة اللغة والأدب، لكي يقوم بعمليات آلية مثل العد والفرز، وضغط الملفات، ومعلومات البحث)) [(3)].

وعرفها «عبد العزيز ميهوبي» تعريفا غامضا منطلقا من غاية جزئية في اللسانيات الحاسوبية بقوله: ((هو دراسة الجوانب الحاسوبية للغة، والمشاكل الشائعة التي تواجه المعالجة الحاسوبية للغة في صورتها المكتوبة أو المنطوقة)) [(4)].

وحاول «مازن الوعر» أن يجد تعريفا لها عبر إيقونة المقارنة مع اللسانيات العامة،

---

(1) مدخل الى اللسانيات الحاسوبية، مجموعة كتاب، 6، مقدمة محرر الكتاب عبد الله بنت يحيى الفيفي.

(2) توظيف اللسانيات الحاسوبية في إعادة توصيف قواعد النحو العربي، إبراهيم إسماعيل المصري، بحث منشور في مجلة كيرالا، العدد 11، جامعة كيرالا، الهند، 2018، ص 265.

(3) دور الذكاء الاصطناعي في تعليمية مستويات اللغة العربية لتلاميذ الطور الابتدائي الحاسوب انموذجا، حمزة بو كثير، بحث منشور ضمن كتاب التكنولوجيا الجديدة ودورها في صناعة اللغة العربية، ص 130.

(4) تطبيقات لسانيات المدونة الحاسوبية على اللغة العربية، مامة كروش، رسالة ماجستير، جامعة بكر بلقايد، كلية الآداب واللغات، 2018م، ص 16.

والشرح نوعا ما بقوله: ((هو الإطار العلمي الذي يبحث في اللغة البشرية كأداة طيعة لمعالجتها في الآلة، وتتألف مبادئ هذا العلم من اللسانيات العامة، بجميع مستوياتها التحليلية الصوتية، والنحوية، والدلالية، ومن علم الحاسبات الالكترونية ومن الذكاء الاصطناعي والمنطق، ثم الرياضيات مشكلة بذلك مبادئ اللسانيات الحاسوبية)) [1].

تنطلق اللسانيات الحاسوبية في مفهومها من الجنبة الرياضية في الأساس، والتعشيق بين العلمين (اللغوي - والحاسوبي)، وإلقاء بظلال المعلوماتية على اللغة، وصار ذلك اللقاء، والتعشيق حتميا؛ ((وذلك لسبب أساسي وبسيط، هو كون اللغة تجسيد لما هو جوهري في الإنسان، أي نشاطه الذهني بكل تجلياته في نفس الوقت الذي يتجه فيه الحاسوب نحو محاكاة بعض وظائف الإنسان، وقدراته الذهنية متخذا من الاعتبارات الإنسانية (الهندسة البشرية) محورا رئيسا لتصميم نظمه ومجالات تطبيقه، ومطالب تشغيله)) [2]، فصار لزاما اللجوء إلى هذه اللسانيات؛ بغية تحديث العربية، والتقدم بها، وطي الصفحة القديمة - وهذا لا يعني هجرانها أو رفضها - حتى تساير العربية الحياة، وتفيد من إمكانيات الحاسوب[3]، وهذا لا يعد قصورا أو تقصيرا في عربية أهل الجنة، أو عجزا في أبحاثها، وقلة جدواها، بل لقابليتها للتطوير، وانفتاحها على الإضافة، ثم لا مناص من تدخل الحاسوب في صغائر الأمور، ومسالك التواصل القائمة على الطباعة، والمحادثة، والتواصل عبر الآلات الذكية، التي تركت إنطباعها على الفرد العربي. فنما هذا الفرع اللساني دامجا اللغة بالحاسوب، وصار يمتلك مفهوما جديدا، ووظيفة جديدة، وتغيرت المفاهيم اللغوية فيه، مثل مفهوم الكفاية اللغوية؛ بسبب عدم استعمال السياق، وتغير مجالات تطبيقه. فمعه تحققت مادة لسانية بحثية مجدولة، أو محدودة، أو مؤطرة لم تتسم بالاتساع والشمول، وإنما بنيت على الطاقة الاستيعابية للحاسوب، أو القدرات التعبوية التي يستقبلها الحاسوب، ويقوم بها الحاسوبيون واللسانيون، فصرنا أمام مُخرج جديد ذي شبكة لسانية جديدة، لن تسع العربية

---

(1) المصدر نفسه.

(2) اللغة العربية والحاسوب، د.نبيل علي: 114.

(3) ينظر: اللغة العربية والحاسوب: 84- 85.

كلها، بناء على توصيف بعضها، ولم تتسم بالتواصلية أو التخمين والحدس، وإنما بناء هاجس الاشتراك بين العلمين على وفق بعض الكليات، التي تسهم في تحليل النصوص القرآنية والأدبية تحليلا لسانيا.

وهذه الكليات تنتمي لعلمين كما هو معلوم، تنتمي إلى اللسانيات العامة، بمستوياتها التحليلية كافة: الصوت، والصرف، والنحو، والدلالة، ثم علم الحاسوب، وعلم الذكاء الاصطناعي، علم المنطق، ثم علم الرياضيات، تأتلف تلك العلوم لتكون النتيجة علم اللسانيات الحاسوبية[1].

يتسم هذا العلم بالآتي:

- بعد إجراء توصيف دقيق للنظام اللغوي يتجاوز المشكلات اللغوية نحو السياق، وظروف إنتاج الكلام أي الميتا لغوي[2].
- يستلزم مراعاة الضبط، والدقة أثناء تخزين البيانات، الإملائية، والنحوية، والصرفية، فلما يغيب الضبط، لم يميز الحاسوب بعض المفردات، ويهيمن اللبس على العملية التواصلية، التي جيء بعلم اللسانيات الحاسوبية من أجلها[3].
- لم يكتف الحاسوب في العملية اللغوية بالخزن والاستدعاء، والتحليل الرقمي لبعض المعطيات اللغوية، وإنما يترجمها إلى لغات أخرى أحيانا[4].
- وما يؤخذ على هذا الفرع اللساني أنه لم يستظهر المعنى ((وليس في ما وصلت إليه أرقى الحواسيب شيء من الملكات العقلية التي يستطيع بها الإنسان تفسير المعنى))[5].
- ينماز بالتشفير، أي تعمية على الملفات، وعدم افتضاحها، لضمان السرية، وخصوصية

---

(1) ينظر: دور الذكاء الاصطناعي في تعليمية مستويات اللغة العربية لتلاميذ الطور الابتدائي الحاسوب انموذجا: 131.

(2) ينظر: دور اللسانيات الحاسوبية في تنمية اللغة العربية وعلومها، بختة حاتي، بحث منشور في المجلة التعليمية، المجلد 5، العدد15، 2018، ص 125.

(3) ينظر: اللسانيات المجال والوظيفة والمنهج: 548.

(4) ينظر: دور الذكاء الاصطناعي في تعليمية مستويات اللغة العربية لتلاميذ الطور الابتدائي الحاسوب انموذجا، بحث منشور ضمن كتاب التكنولوجيا الجديدة ودورها في صناعة اللغة العربية: 130.

(5) اللسانيات المجال والوظيفة والمنهج: 569.

الأفراد، وأنشطتهم[1].

خرجت اللسانيات الحاسوبية إلى الفضاء اللساني في الفكر الأوربي قبل العربي بمـدة طويلة، متزامنة مع الحرب العالمية الثانية، عبر طريقة الترجمة بواسـطة آلـة (إنجمـا) Enigma، ويقسم الدكتور نبيل علي تلك النشأة على ثلاث مراحل[2]:

المرحلة الأولى: الارهاصات، مع مرحلة اكتشاف الحاسوب نفسه، ومـسألة احتياجـه إلى اللغة[3]، ومعالجة البيانات، وهذا في بداية الأربعينيات من القرن العشرين.

المرحلة الثانية: البداية، في نهاية الاربعينيات مـن القـرن العـشرين، اذ وظفتـه بعـض السياسات توظيفا عسكريا، وأمنيا في تحليل بعض البيانات، وترجمتهـا، للإفـادة منهـا في إدارة عمل المخابرات.

المرحلة الثالثة: الازدهار، في منتصف القرن العشرين، على يد جومـسكي، إذ وضـع أسس الكتابة الشكلية للنحو، والجنبة الرياضية فيه، والأساس البيولوجي، والأساس النفـسي للغة، فصارت تلك النظرات أساسا، وتمهيدا للسانيات الحاسوبية.

تلك اللسانيات التي كانت الترجمة العامـل الأسـاس في نـشوئها، إذ حـددت بـدايتها بعام 1949م حينما كتب وارن ويفر مذكرته التي أشاد فيها بإمكانية بناء نظام الترجمـة الآلـي، فعُقد بناء على ذلك مؤتمر الترجمة الآلي في معهـد (ماساتـشوستس للتكنولوجيـا) (MIT) في عام 1952 ، وبعدها بعامين صدرت مجلة الترجمة الآلية[4].

ويقرب الدكتور «نبيل علي» نشأتها من شانون، ونظريته نظريـة المعلومـات، وأثرهـا في سلاسل ماركوف الإحـصائية، الـتي وضـعت أسـاس النمـوذج الإحـصائي للغـة، وصـار بالإمكان الإفادة من معيـار دقيـق لقيـاس المعلومـات اللغويـة وكميتهـا، ثـم جـاءت مرحلـة

---

(1) ينظر: اللغة العربية والحاسوب: 241.

(2) ينظر: تطبيقات لسانيات المدونة الحاسوبية على اللغة العربية: 21 وما بعدها.

(3) هنالك اثر متبادل بين الحاسوب واللغة، فكما ان اللغة تحتاج الحاسوب في تحليلاتها، ودراساتها، ومعالجاتها، يحتاج الحاسوب اللغة، اذ لا يستقيم بدونها ولا يتحقق الفهم، أصلا .

(4) ينظر: مدخل الى اللسانيات الحاسوبية: 5- 6، مقدمة الكتاب بقلم عبد الله يحيى الفيفي.

التحليل اللغوي، والترجمة الآلية، التي تلتها نظرية النموذج الرياضـي لجومـسكي[1]، فاتكـأت اللسانيات الحاسوبية على تلك التكوينات، وبدأت بالصدور، والظهور، وتنازعتهـا تـسميات متعددة، الآلية اللغوية والحسابية اللغوية وعلم اللغة الحاسبي، وعلم اللغـة المحوسـب، وعلـوم اللغة الحاسوبية، واللغويات الحاسوبية، واللسانيات الحسابية، والهندسـة اللـسانية، والألـسنية المعلوماتية، والهندسة التقنيـة للغـة، والهندسـة التكنولوجيـة للغـة[2]. كـان الأكثـر اسـتعمالا وشيوعا مصطلح اللسانيات الحاسوبية، وبـدأ ظهـوره في منتـصف الـستينيات، إذ إن «ديفيـد هيز» أول من أطلق هذا المسمى يوم كان عضوا في اللجنـة الاستـشارية لمعالجـة اللغـة آليـا في الأكاديمية الوطنية للعلوم في امريكا[3].

ثم جاءت الجهود العربية بعد أن عُقدت الندوات في دول المغـرب العربـي ولاسـيما النـدوة الدوليـة لجمعيـة اللـسانيات في الربـاط في 12 ابريـل 1987وجـاءت جهـود العـرب بعضها على شكل أبحاث مثل محاولات عبد الرحمن الحاج صالح، ومنهـا بحثـه اللغـة العربيـة من منطق الخليل إلى الحوسبة الخليلية، إذ عرض إفادة العربية من التكنولوجيا، وجـاء بعـضها على شكل كتب مؤلفة مثل كتاب الدكتور نبيل على اللغة العربية والحاسوب، طرح فيـه كـل ما يتعلق بهذه النظرية، وكتاب نهاد الموسى توصيف العربية حاسـوبيا، ثـم الاتجـاه التطبيقـي على اللغة العربية بدراسـات أكاديميـة ومنهـا دراسـة الباحـث باسـم صـالح النـداء في شـعر الجواهري دراسة في ضوء اللسانيات الحاسوبية.

وثمة دوافع على نشوء هذا الفرع اللساني، وقـد قـالوا بقـصور الدراسـات اللغويـة السابقة، ولست مع هذا الرأي الجارح، والمصادر لنظريات كثيرة. وإنما هنالك مـشكلة تعلـيم العربية، ومدى تعلق الناس بالحاسوب، فـصار لـه أثـر في تجـاوز الآثـار الـسلبية في التعلـيم، والـسبب الثالــث دخــول بعــض تطبيقــات الحاســوب في مجــالات الإنــسانيات والعلــوم الاجتماعية، والسبب الرابع الافتراء على العربية وعدم تقبلها للمعالجـة الاليـة، فـسعى هـذا

(1) ينظر: اللغة العربية والحاسوب: 73- 74.

(2) ينظر: هذه المصطلحات المتعددة في : تطبيقات لسانيات المدونة الحاسوبية على اللغة العربية: 17.

(3) ينظر: مدخل الى اللسانيات الحاسوبية: 6، مقدمة محرر الكتاب عبد الله يحيى الفيفي.

التوجه الى تبديد هذا الخطر[1]. فضلا عن إنفجار المعلومات الهائل، وإنعكاسه على العربية، والميل نحو اختصار الوقت والجهد في معالجة العربية.

وهنالك أهداف من دراسة هذا الفرع[2]:

- يهدف إلى رصد الظواهر اللغوية بمستوياتها الصوتية والصرفية، والنحوية، والدلالية، والبلاغية، والعروضية.
- يهدف إلى معالجة التحليل اللغوي للنصوص، عبر الإحصاءات.
- يهدف إلى صناعة المعجم الرقمي.
- يهدف إلى تسهيل عملية التعليم للغات.
- يهدف إلى اختصار الوقت، وتحقيق السرعة في البحث، ولا سيما في سرعة الترجمة، وجعلها فورية.

أدرك اللسانيون مجالات للسانيات الحاسوبية، مستوحاة من اللسانيات الصُورية أو مما قبل البينية:

- **المجال الصوتي:**

ويقوم على تمييز الكلام، وتوليده آليا، بناء على إشارات طيفية يستوعبها الحاسوب، بطرائق هندسية ورياضية، تفرز السمات الأساسية للاصوات المنطوقة، بما فيها الكلمات المنعزلة، والكلام غير المتصل، والكلام المستمر (المتدفق) [3].

- **المجال الصرفي:**

تحلل الكلمات في الحاسوب صرفيا بعد ان تستوعَب المتعلقات الصرفية من الجذور، والجذوع، والسوابق، واللواحق، وكل ما يتعلق بالميزان الصرفي، والاشتقاق، وبعض الحالات الاعرابية[4]، وهنا يعاني الحاسوب من صعوبة المجال الصرفي الحاسوبي في اللغة

---

(1) ينظر: اللغة العربية والحاسوب: 8- 10.

(2) ينظر: تطبيقات لسانيات المدونة الحاسوبية على اللغة العربية: 13.

(3) ينظر: اللغة العربية والحاسوب: 136.

(4) ينظر: اللغة العربية والحاسوب: 138- 139

العربية قياسا باللغات الاوربية؛ كونها لغة اشتقاقية، تقوم على نظام الجذر اللغوي، والاوزان الصرفية[1]، ويتوقف على هذا المجال اغلب مجالات التحليل الآلي[2].

- **المجال النحوي:**

يتم توصيف المادة النحوية، وتستدرج في الالة على شكل برمجيات، وايعازات[3]، فيحلل الحاسوب علاقة الكلمات بعضها ببعض، بعد الاعتماد على المجال الصرفي، فيبدأ العمل القواعدي، مجردا من السياق[4].

- **المجال المعجمي:**

اذ يتم فيه برمجة الحاسوب، واستقبال كم كبير من المفردات، ومعانيها، وما يتعلق بها من علاقات دلالية، مثل الترادف او المشترك اللغوي، او الاضداد، وبعض السمات الصرفية للمفردة.

- **المجال الدلالي:**

ويعتني بفهم الدلالات الكامنة وراء الجمل، والترابط المنطقي بينها، ويفيد في الوقت نفسه من المجالات السابقة ولاسيما المجال الصرفي، والنحوي، والمعجمي[5].

- **مجال الترجمة:**

وهو المجال الأرحب، والأهم في اللسانيات الحاسوبية، وتصب في نظم التحليل، والتركيب اللغويين، كونها تهيمن على المنظومة اللغوية التقابلية[6]، فتشتغل في هذا المجال الحاسوبي الترجمة الالية التي ((تعني استخدام برمجيات الحاسب في ترجمة النصوص من لغة إنسانية الى لغة إنسانية أخرى)) [7].

---

(1) ينظر: معالجة اللغة العربية اليا دراسة حاسوبية ، عمرو جمعة: 23.

(2) ينظر: مدخل الى اللسانيات الحاسوبية: 47- 48.

(3) ينظر: العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية: 79.

(4) ينظر: دور الذكاء الاصطناعي في تعليمية مستويات اللغة العربية لتلاميذ الطور الابتدائي الحاسوب انموذجا: 133.

(5) ينظر: دور الذكاء الاصطناعي في تعليمية مستويات اللغة العربية لتلاميذ الطور الابتدائي الحاسوب انموذجا: 134.

(6) ينظر: اللغة العربية والحاسوب: 145.

(7) معالجة اللغة العربية اليا دراسة حاسوبية لظاهرة اللبس اللغوي في العربية المعاصرة: 124

- **مجال التعليم اللغوي:**

يدخل الحاسوب في مجالات التعليم كلها، ومنها تعليم اللغات، ((فقد اتسع استعماله ليشمل العملية التعليمية التعلمية، حيث أصبحت تتعدد مجالات استخدامه سواء كهدف تعليمي أم كأداة أم كعامل مساعد أثناء العملية التعليمية أم كمساعد في الإدارة التعليمية)) [1]، ويترتب على وجود الحاسوب في التعليم سرعة الأداء، والدقة، والانجذاب، والتجدد، وتجاوز البوتقات القديمة.

- **مجال الكتابة:**

وفيه يجري التدقيق الإملائي، والصرفي، والنحوي، عبر وضع برامج وظيفتها التدقيق، وإصلاح الزلل، بعد تخزينها في الحاسوب بشرط صوابها، فضلا عن ذلك التدقيق يدخل في مجال الكتابة التعرف على الكلام المكتوب، وتحويله إلى مادة صوتية، أو على العكس[2].

- **مجال التحليل النصي:**

أي تحليل النتاجات الأدبية، سواء أكان شعرا أم نثرا، وبيان الخصال الجمالية فيها، وتحديد موضوعاتها، عبر المؤشرات الإحصائية في المفردات، والصيغ الصرفية، وبعض الأساليب النحوية، وأدوات الفصل والوصل، والميزات النحوية الأخرى، بوصفها أدوات ربط[3].

- **المجال التواصلي اليومي:**

لم يلتفت إلى هذا الجانب إلا الدكتورة آمنة بلعلى وأطلقت عليه (الشفوية الجديدة)، بمعنى تخلي الناس عن التواصل عبر الكتابة الورقية، وتشبثهم بالكتابة الالكترونية الشفوية

---

(1) دور الذكاء الاصطناعي في تعليمية مستويات اللغة العربية لتلاميذ الطور الابتدائي الحاسوب انموذجا: 122

(2) ينظر: اللسانيات الحاسوبية العربية المفهوم المنطلقات الجدوى: 74.

(3) ينظر: اللغة العربية والحاسوب: 142- 144.

عبر الرقميات، والهواتف النقالة، مواقع التواصل[1].

- **مجال الأسلوبية الحاسوبية:**

حين استوى عود اللسانيات الحاسوبية، وتجاوزت مرحلة التأسيس، بدأت تتفرع شأنها شأن اللسانيات العامة، وتسير الى جنبها، فانفتحت على مناهج تحليل الخطاب، وصار ما يعرف بالأسلوبية الحاسوبية، باستخدام الحاسوب في المقدرة على التحليل، والإحصاء المفصل للنصوص[2].

هذه المجالات اللسانية الحاسوبية ، بعضها نصي: يتعلق بالجانب الدلالي، والتحليل النصي، وبعضها خارج نصي، يتعلق بدراسة مستويات اللغة التي عرفناها في علم اللغة العام، من قبيل المفردات، والصيغ الصرفية، والجزئيات النحوية، وبعضها خطابي يعنى بتحليل الخطاب اذا فهمنا الخطاب على أنّه التواصل والتحادث بين المتكلمين، وبعضها خارج هذه الإيقونات كلها، ينتمي إلى مجال الترجمة، ومجال التعليم اللساني، فالترجمة علم مستقل وان انتمى الى علم اللغة التقابلي في ضفة ما، وقد يحسب هذا التوظيف الجديد، مثل ما يحسب لها جدة مجال التعليم، وهذا يؤشر التداخل بين اللسانيات التربوية واللسانيات الحاسوبية، وافادة بعضمها من بعض، مع هذه الملاحظات يبرز مخطط لتطبيقات او لمجالات اللسانيات الحاسوبية يتمثل في الاتي:

1- التحليل والتركيب الالي: الترجمة الالية.

2- التدقيق الآلي:

أ- تصحيح الاملاء.

ب- التصحيح النحوي.

3- التشكيل الآلي:

أ- التلخيص الالي.

---

(1) ينظر: التكنولوجيات الجديدة وتسويق الشفوية الجديدة، امنة بلعلى، بحث منشور ضمن كتاب التكنولوجيات الجديدة ودورها في صناعة اللغة العربية: 31.

(2) ينظر: دور الذكاء الاصطناعي في تعليمية مستويات اللغة العربية لتلاميذ الطور الابتدائي الحاسوب انموذجا:136.

ب- التعرف الالي على الكلام.

ج- التعرف الالي على الخط المكتوب.

4- التركيب الآلي:

أ- توليد آلي للكلام.

ب- توليد آلي للمعجم.

ولم يستقيم هذا العلم الجديد، ولم يستقر اذا ما تولد معه مفاهيم جديدة، مثل ما انبرى منهج جديد لدراسة اللغة، فبقيت تلك المفاهيم بمحاذاة المفاهيم اللسانية العامة.

منها مفهوم الكفاية اللغوية، فأخذ يدل على كمية استدخال قواعد اللغة في أنظمتها كافة من جهة الصرف، والصوت، والنحو، والاملاء، والمفردات، ووجوه الاستعمال، وكذلك يدل على قيمة الإنتاج من الأداءات اللغوية الصحيحة، فصارت دالة على انها كفاية تواصلية[1]، أو أدائية كما يسميها نهاد الموسى ((تناظر كفاية الانسان العربي العارف بلغته، القادر على تمييز المكتوب والمنطوق، وتأليفهما، القادر على توليد الأبنية وتحليلها وتمييزها، وتأليف التراكيب وتحليلها، وتبيين الأعاريب وأجرائها، وفهم دلالات الألفاظ ووجوه استعمالها، وتصحيح الإملاء)) [2].

فيصير مفهوم اللغة في اللسانيات الحاسوبية تلقي المعلومات وارسالها، أو تقديمها بلا تحوير أو إضافة او تغيير، مجردة من السياق تماما، وإنما يتم التواصل عبر اركان العملية التواصلية في محورين مهمين.

المحور الأول محور الترجمة، فيكون أركان عملية التواصل في حالة الترجمة:

1- المتكلم الفرد.

2- المتلقي الفرد.

3- الأثر المترجَم، او المقروء (النص).

---

(1) ينظر: اللسانيات الحاسوبية المفهوم التطبيقات الجدوى: 63- 64.

(2) العربية نحو توصيف جديد دراسة في ضوء اللسانيات الحاسوبية: 34.

المحور الآخر: محور القراءة، فيكون أركان عملية التواصل فيها:

1- المتلقي الفرد.
2- الأثر النص.
3- القناة الشاشة.

يتضح ثمة فرق شاسع بين اللسانيات العامة، واللسانيات الحاسوبية، وما يميزها أكثر تلك السمات البينية، التي أقف عليها:

- **التخزين:**

بولادة الحاسوب، وتلاقحه مع اللغة، وتاثر أحدهما بالآخر صار لزاما على البيانات أن تدخل الى الحاسوب، وابرازها على الشاشة، وهنا مرحلة التخزين جاءت بديلة عن النظام الكتابي في اللسانيات العامة، اذ يقوم الكاتب أو الباحث بتسجيل النص المراد تحليله، ثم اجراء التطبيق عليه، بينما في اللسانيات الحاسوبية على العكس من ذلك، اذ تخزن البيانات ويضغط احد الايعازات الى الحاسوب، ويجري التحليل[1].

- **التدوين اللغوي (المدونة اللغوية):**

المدونة اللغوية جسم من المادة المكتوبة او المنطوقة، يتم جمعها من جهات مختلفة، كتب او صحف، او لقاءات إذاعية، وتلفزيونية، ودوريات، وغير وتنسق تنسيقا سليما، وتتم فهرستها، ويجري تخزينها في الحاسوب؛ كي ينتفع منها المستعملون او المحللون اللغويون، البينية فيها كونها تتلاقى المادة اللسانية فيها مع الحاسوب على شكل مجموعة خاصة[2]، فلا توجد لولا وجود الحاسوب، بل هي وجدت معه.

---

(1) ينظر: اللغة العربية والحاسوب: 232.

(2) ينظر: تطبيقات لسانيات المدونة الحاسوبية على اللغة العربية: 10.

- **البنكية:**

البنكية من نقاط البينية الأخرى التي برزت ببروز الحاسوب، وتعني مستودع يحتفظ بالمعلومات اللغوية التي تنتج خلال سير النظرية اللسانية، لتكون منبعا للباحث في اللغة، ولكثير من الدراسات في العلوم الأخرى، التي تشتغل عليها بعض النظريات في التاريخ والعلوم الاجتماعية[1]. وبدأت تنشطر عن مفهوم البنكية مفاهيم أخرى تحمل التوصيف نفسه، مثل البنك الشجري treebank ويقال فيه: ((بانه مدونة لغوية توصيفية، تتضمن بعض التحليلات النحوية التي تتجاوز بيان مستوى الأجزاء الكلامية الى الكشف عن العلاقات التركيبية بين هذه الأجزاء)) [2].

- **الآلية:**

الآلة من مفاهيم الحاسوب، بل الحاسوب نفسه آلة، وفيه آلات أخرى، فعند التمازج أو عند المزج، صارت الآلية من أهم النقاط البينية بين اللسانيات والحاسوب؛ إذ أفادت اللسانيات كثيرا من هذا المنحى، واستغلت خدمة الحاسوب، وطواعيته في التحليل الألسني، وطبقته على مستويات اللغة كافة، فكان التحليل الصرفي، والتحليل النحوي، والتحليل الدلالي، والتحليل النصي، ووجدت نقطة بينية جديدة تسمى التلخيص الآلي، وأيضا المُشكِّل الآلي، والترجمة الآلية، واخرها التدقيق الآلي.

يتوقف التحليل الآلي وأنواعه على مستوى بنكية اللغة، فعد إيداع المادة اللسانية في مستودعات حاسوبية، تحلل تلك المادة صرفيا فيسمى التحليل الصرفي الآلي، القائم على ((ربط كلمات النص بالعناصر الصرفية الأولية التي تدخل في تكوينها، وكذلك بالقيم النحوية دون اعتبار موقعها)) [3]، فيتم في الكلمة العودة الى جذرها الأصلي، وبيان جذعها، ثم التغيرات التي طرأت عليها من اعلال وابدال وادغام وزيادة ونقص وبيان الميزان الصرفي

---

(1) ينظر: بحوث وراسات في اللسانيات العربية، عبد الرحمن الحاج صالح:2/ 155.

(2) البنك الشجري النحوي حاسوبيا: بناؤه وتوظيفه في إطار تقنيات الذكاء الاصطناعي، أحمد روبي محمد 11.

(3) مدخل الى اللسانيات الحاسوبية: 50.

وبعض السوابق واللواحق، والذي يقوم بذلك هو المحلل الصرفي الآلي، عبارة عن تطبيق حاسوبي يبرمج على التصورات الصرفية مضافا إليها المادة الصوتية والنحوية[1].

والذي يقوم بالتحليل النحوي يسمى المحلل النحوي الآلي، إذ ((يقوم المحلل النحوي الآلي بتفكيك الجمل إلى عناصرها الأولية، أو بعبارة أخرى يحلل الجملة إعرابيا، ويستخلص العلاقات النحوية المختلفة، تمهيدا لتمثيل بنية الجملة بصورة تفصيلية سافرة))[2]. وثمة مشكلة هنا تتمثل في التغيرات التي تطرأ على الجملة؛ إذ تُحدِث لبسا في التحليل، فيفوت الحاسوب إدراك التقديم والتأخير ، والحذف، والاستتار وغيره في الجمل[3].

وفي الجانب النحوي يستلزم استيعاب أو استقبال التنميط النحوي الذي سماه نهاد الموسى (العيار)، بحيث نضع الأعيرة النحوية في النظام الحاسوبي، وعلى أساسها يبدأ التحليل النحوي من تلقاء الحاسوب نفسه، لتتمثل عندنا خطوة بينية واضحة.

وبعد أن تكون هنالك قاعدة بيانان معجمية، ولغوية بمستوياتها المعروفة، يتم تحديد معاني الكلمات منفردة عبر وسائط مثل شبكة الكلمات، ويتم تجميعها لتحديد المعنى الإجمالي للجملة، وتسمى هذه العملية التي يقوم بها المحلل الالي التحليل الدلالي الالي[4]، ليفك اللبس الدلالي في النص، وليحلل المشاعر في مواد النص.

ويعرف تحليل النصوص على الحاسبة بالتحليل الآلي للنصوص، عبر معرفة السمات الصرفية، والنحوية، والمفرداتية، مضافا إليها المزاج العام[5]، والمعلومات الدلالية، وتنتزع المعلومات بتطبيق مناهج إحصائية متقدمة[6].

---

(1) ينظر: معالجة اللغة العربية اليا دراسة حاسوبية لظاهرة اللبس اللغوي في العربية المعاصرة: 25.

(2) اللغة العربية والحاسوب: 139.

(3) ينظر: اللغة العربية والحاسوب: 140.

(4) ينظر: مدخل الى اللسانيات الحاسوبية: 111.

(5) بمعنى البنية للكلية للنص او الدلالة الشاملة له، تستنبط من خلال التمعن بالنص وتأمله، بناء على الإحساس والتأويل الداخلي للقارئ.

(6) ينظر: مدخل الى اللسانيات الحاسوبية: 130.

ومن التطبيقات اللغوية الحاسوبية التي تندرج في الالية (التلخيص الآلي) بعد أن يتم إنشاء نص في ملف او في مستند الي، ثم يتم اختصاره اليا مع المحافظة على اهم أفكار النص الأصلي.

ويأتي المشكل الالي بعد ذلك ليشكل النصوص الملقاة على عاتق الحاسوب أو على شاشته، او يشكل اهم الكلمات ذات اللبس بحسب البرامج المعدة لذلك.

ومخرج الترجمة الآلية من أهم مخرجات اللسانيات الحاسوبية في العصر المعلوماتي، وتتم فيه ترجمة النصوص من دون تدخل الفرد، وتحول الكلام من لغة الى لغة أخرى، ولها توظيفات عدة في المجال التعليمي، والتجاري، والصناعي، وغير ذلك.

وينتمي التدقيق الآلي الى جنبة الكتابة، والطباعة، والاملاء، اذ يؤشر الحاسوب - بعد أن يزود ببرنامج خاص بذلك- الخلل الحاصل في بنية الكلمة الاملائية، والنحوية، فكان عندنا نوعان من التدقيق، التدقيق الإملائي، والتدقيق النحوي، اذ يعطي خيارات عدة صحيحة، ويختار احدها القائم بأعمال التدقيق، ويصدف أن لا يوجد الرسم الصواب، فيجري القائم بالتدقيق تصحيحا ويضيفه غلى قاموس الحاسوب.

- **المكننة:**

مصطلح المكننة من المصطلحات الصناعية أكثر منه لغويا، وبفضل الممازجة صار يوظف توظيفا بينا، واختص في العرف اللساني بالعجم، ونتج ما يسمى بالمعجم الآلي، أو المعجم الممكنن، أو المعجم الرقمي، أو القاموس الرقمي، أو معجم عصر المعلوماتية، وهو عبارة عن إدخال بيانات اللفظة على أساس معين، كأن يكون الجذر مثلا، وتوجد بعض المبادئ في التنظيم المعجمي الممكن[1]:

1- لابد من تخزين كل المتعلقات الافرادية، واللفظية، والتركيبية: أصولها وفروعها، نحو الأوزان، وما يكثر من تلك الأوزان ولاسيما في العربية قرابة 300وزن، وكذلك المثل اللفظية، والتشكيلات التركيبية.

---

(1) ينظر: بحوث ودراسات في اللسانيات العربية: 1/ 93- 94.

2- فرز الأوزان الخاصة بالاسم، والخاصة بالفعل، وما يشترك بينهما، وفرز اوزان الاسم المفردة، والمثنى، والجمع.
3- الكشف عن جنس الكلم ومدلولاتها، وجنس الجمل أيضا، وعامة التراكيب.
4- خزن المفردات كثيرة الدوران، والشائعة، بدلالتها الوضعية، ودلالتها الفرعية، حتى نتتفع من ذلك الكم المعجمي في عملية التحليل اللفظي.

- **قاعدة البيانات:**

من اهم المفاهيم في علم الحاسوب، ولما أفادت منها اللسانيات الحاسوبية صار ما يعرف بقاعدة بينات نحوية، وقاعدة بيانات معجمية، وهكذا اذ تسهل الرجوع إلى المواد اللسانية عبر طريقة الرسيرج (البحث).

- **التعرف:**

التعرف على بصمة الفرد، وكشف الأصوات كان يتم في اللسانيات العامة عن طريق الأذن، وبعض المختبرات الصوتية، ولما توفر الحاسوب كان له أثر في التعرف آليا على بعض الأصوات بناء على برمجيات، وتخزين للأصوات المعنية، فـ((التعرف الآلي على الكلام هو تطوير نظام حاسوبي، يتولى تحويل الموجة الصوتية الى نص)) [1]، بعد تعرفه على الخصائص الاكوستيكية، ومقارنتها بما مخزون فيه[2].

- **التعبيرية:**

بينما كان التعبير عن طريق الجهاز النطقي عند الإنسان، صار التعبير، وتوليد الكلام عن طريق الحاسوب آليا، اذ يتم تحويل النص المكتوب إلى موجات مسموعة، واضحة[3].

---

(1) مدخل الى اللسانيات الحاسوبية: 28.
(2) ينظر: مدخل الى اللسانيات الحاسوبية: 28.
(3) ينظر: مدخل الى اللسانيات الحاسوبية: 30.

الخاتمة

بناء على ما مرَّ من عرض موجز للسانيات الحاسوبية تبين مدى تأثر اللغة بالحاسوب، وتأثر الحاسوب باللغة في علاقة متبادلة مبنية على التأثير والتأثر، وهذه العلاقة أنتجت لنا وجودات لسانية جديدة، تخدم البحث اللساني، وتتجاوز كثيرا من عقباته التي يمر بها عبر المنهج الاحصائي او المعرفي او الخوارزمي كما يصنفون تلك المناهج في هذه اللسانيات، وتوفر مساحة حداثوية تواكب التطور التكنولوجي في العصر الحديث، وظهرت عندنا بوادر بحثية وتواصلية اذا لم تتوفر الحاسبة لم تتوفر تلك المخرجات من قبيل التحليل الآلي أو الممكنة وغيره مما ينتمي الى هذا المجال الالكتروني او الحاسوبي، ففرضت طريقة تعبير جديدة، وطريقة تكنيك بحثية مبتكرة في تحليل النصوص والغوص في دلالتها، وميزاتها الأسلوبية بناء على الإحصاء اللغوي بعد هذه الإحصائية هي الركيزة الأساس في هذا العلم اللساني الحاسوبي، يضاف الى ايقونة الإحصاء بعض التصورات اللسانية العامة ويتم توصيفها وتخزينها في الحاسوب عبر اجسام خاصة تسمى الذخائر اللغوية، لتكون مادة سهلى عند مراجعتها ودراستها على وفق اللسانيات الحاسوبية.

المصادر

- توظيف اللسانيات الحاسوبية في إعادة توصيف قواعد النحو العربي، إبراهيم إسماعيل المصري، بحث منشور في مجلة كيرالا، العدد 11، جامعة كيرالا، الهند، 2018.
- العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، نهاد الموسى
- اللسانيات الحاسوبية العربية المفهوم التطبيقات الجدوى، وليد أحمد العناتي، مجلة الزرقاء للبحوث والدراسات، المجلد 7، العدد الثاني 2005
- مدخل الى اللسانيات الحاسوبية، مجموعة كتاب، مقدمة محرر الكتاب عبد الله بنت يحيى الفيفي.
- دور الذكاء الاصطناعي في تعليمية مستويات اللغة العربية لتلاميذ الطور الابتدائي الحاسوب انموذجا، حمزة بو كثير، بحث منشور ضمن كتاب التكنولوجيا الجديدة ودورها في صناعة اللغة العربية،
- تطبيقات لسانيات المدونة الحاسوبية على اللغة العربية، مامة كروش، رسالة ماجستير، جامعة بكر بلقايد، كلية الآداب واللغات، 2018م،
- اللغة العربية والحاسوب، د.نبيل علي:
- دور اللسانيات الحاسوبية في تنمية اللغة العربية وعلومها، بختة حاتي، بحث منشور في المجلة التعليمية، المجلد 5، العدد15، 2018،
- اللسانيات المجال والوظيفة والمنهج
- تطبيقات لسانيات المدونة الحاسوبية على اللغة العربية:
- معالجة اللغة العربية اليا دراسة حاسوبية لظاهرة اللبس اللغوي في العربية المعاصرة:
- اللسانيات الحاسوبية العربية المفهوم المنطلقات الجدوى:

- التكنولوجيات الجديدة وتسويق الشفوية الجديدة، امنة بلعلى، بحث منشور ضمن كتاب التكنولوجيات الجديدة ودورها في صناعة اللغة العربية:
- بحوث وراسات في اللسانيات العربية، عبد الرحمن الحاج صالح: 2/ 155.
- البنك الشجري النحوي حاسوبيا: بناؤه وتوظيفه في إطار تقنيات الذكاء الاصطناعي، أحمد روبي محمد
- معالجة اللغة العربية اليا دراسة حاسوبية لظاهرة اللبس اللغوي في العربية المعاصرة.

المعالجة الآلية للمستوى الصوتي في اللسانيات الحديثة

أ.م. د. بشرى عبد الرزاق العذاري
كلية التربية / المستنصرية

المقدِّمة:

الحمد لله رب العالمين، والصلاة والسلام على أشرف الأنبياء والمرسلين، وبعد:

فإن علم اللغة الحديث يَعد الأصوات الإطار الذي تبنى عليه البنى الأخرى، فهي أساس تأليف الكلمات غير المتناهية، ودراستها تعد جزءًا لا يتجزأ من المستويات اللسانية للنص اللغوي، وتقدمت الدراسات الصوتية في هذا العصر تقدما كبيرا، وأصبح الأصواتيون يستعينون بأجهزة دقيقة، سواء في تسجيل الصوت أو في تحليله، مما دعا لظهور فرع جديد من فروع علم الأصوات اصطلح عليه (علم الصوتيات المعملية أو التجريبية) الذي يُعرف بأنه: استخدام الأجهزة والآلات والأدوات الفنية في الدراسة الصوتية في مكان يُعد لذلك، يسمى المختبر الصوتي أو معمل الأصوات، وتطورت هذه الآلات وأصبحت أكثر دقة وأصغر حجما، وأيسر استخداما، ثم انتشرت برامج الحاسوب في مختلف مجالات المعرفة، وامتدت لتشمل معالجة اللغات الطبيعية أو ما يعرف ببرامج حوسبة اللغة، وهي البرامج التي تكسب الحاسوب قدرة على التعامل مع لغة ما بما يشبه تعامل العقل البشري، وذلك إذا امتلك قاعدة بيانات جامعة مانعة – قدر الإمكان- للغة التي يتعامل معها، إلى أن آلت بأن تكون علم مستقل، هوعلم اللغة الحاسوبي أو اللسانيات الحاسوبية Computational Lingustic، أحد فروع علم اللغة التطبيقي، الذي يقوم بتطبيق نتائج المنهج اللغوي وأساليبه الفنية في التحليل والبحث في ميدان غير لغوي.

وفي هذا الإطار، حاول هذا العلم ربط علاقة بين علمي اللسانيات والمعلوماتية، قصد معالجة اللغات الطبيعية معالجة آلية، إذ تتألف مبادئ هذا العلم من اللسانيات العامة بجميع مستوياتها التحليلية : الصوتية، والنحوية، والدلالية.

فمعالجة اللغة العربية حاسوبياً أصبحت اليوم أمراً لا مفرّ منه، ولا سيما أن استثمار الدراسة الحاسوبية والمعلوماتية -بصفة عامة- يحقق نتائج كبيرة للغة العربية، في مجال التعريب، والإحصاء اللغوي، والمعالجة الآلية، وتعلم اللغات، والترجمة الآلية، وفي مجال التربية والتعليم، وهذه الاستفادة تزداد يومًا بعد يوم؛ وبذلك حل الحاسوب محل معامل التحليل الصوتي بحيث يقوم الحاسوب بفحص الأصوات وتصنيفها وغير ذلك من خلال تحويل الإشارات الصوتية المسموعة إلى صورة مرئية توضح الخصائص الأكوستيكي المختلفة (التردد، والشدة، والزمن)، ومعالجتها آلياً بوساطة تحليل طيف الصوت، وتوليد (إنتاج) الكلام، وتخزين الأنماط الصوتية للشخص المتكلم، وتبعاً لهذا صُممت أجهزة تخليق الكلام وتحليله، وتوليد الكلام المنطوق آلياً بتحويل النصوص المدخلة في جهاز الحاسوب إلى مقابلها الصوتي، وعلاج عيوب النطق، وأنجزت دراساتٌ عدة في هذا المجال.

وجاءت وريقاتي هذه، لتقف على آلية المعالجة الحاسوبية الحديثة في التحليل الصوتي للنص اللغوي، وقد قسمت هذه الوريقات على أربع نقاط، هي:

أولًا: المناهج اللغوية المستعملة في المعالجة الآلية للمستوى الصوتي.

ثانيًا: مبادئ المعالجة الآلية الحاسوبية للمستوى الصوتي.

ثالثًا: الصعوبات التي تواجه الباحث في المعالجة الآلية للأصوات.

رابعًا: أشهر البرامج الحاسوبية المستعملة في المعالجة الآلية للأصوات.

وقد سبقتها بمقدمة وانهيتها بخاتمة .

أولًا: المناهج اللغوية المستعملة في المعالجة الآلية للمستوى الصوتي:

فإن اللسانيات الحاسوبية لها جانبانِ تُدرس من خلالهما: الجانب النظري، والجانب العملي، "فالجانب النظري: يهتمُّ بتوصيف هذا الفرع من اللسانيات، ووضْع نظرياتٍ وقواعدَ تُسَاهِم في تحقيق غاياته التي وُجِدَ مِن أجْلها. أما الجانب العملي: فيُطبِّق هذه النظريات والقواعد على اللغة، ويربطها بالحاسب، ويُظهِر مَدى اطِّرَادِها وانضِباطِها، ويُساهِم أيضًا في اكتِشافِ ثغراتها؛ مِن أجْل تطويرها وتعديلها؛ مِن أجْل الوصول إلى أفضل النتائج.

ويلاحظ على أغلب البحوث أنها انتقلت باللسانيات الحاسوبية من مجالها النظري أو التنظيري إلى الجانب التطبيقي، ومن المعروف أن الجانب التطبيقي، هو الجانب الأهم في اللغويات الحاسوبية لأنه يتمثل في تسخير العقل الإلكتروني لحل القضايا اللغوية، وهنا يبرز الدور الرئيس والأثر الفاعل لالتقاء اللغويين والحاسوبيين، والتعاون فيما بينهم، وما يثمر عنه من نتائج تسهم إلى حد كبير في تذليل العقبات وحل المشكلات التي تواجه التحليل الحاسوبي للغة العربية.

وفي ظل التسهيلات العلمية التي قدمتها التكنولوجيا في مجال استخدام أجهزة القياس الصوتية، وإعداد البرامج التي من شأنها إنجاز الكثير من المعالجات الرقمية الآلية السريعة لبيانات المعلومات الصوتية التي يتم تغذيتها لهذه البرامج من خلال الحاسب الآلي، الأمر الذي لا يترك لنا سوى تحديد اختياراتنا بشأن تعيين القيم الرقمية التي نريد أن يستخلصها لنا برنامج صوتي ما، وتحديد أي من الملامح الصوتية التي نسعى إلى دراستها في النص اللغوي، سواء كانت قيم تردد الأساس أو معدلات الشدة الصوتية، أو الضغط، أو الاستغراق الزمني، أو الوقوف على سمات بعض القطوع الصوتية، سنجد أن ذلك كله أصبح متاحًا، ولكن يبقى دائما دور الباحث اللغوي في وصف وترجمة هذه البيانات والمعلومات الرقمية المعطاة وتحليلاتها الطيفية إلى خصائص لغوية تؤدي دورًا وظيفيًا فعالًا في الجانب الذي تشغله هذه الظواهر من النظام الفونولوجي بلغة ما.

وبناءً على ما سبق، فلا يمكن الاستغناء عن المنهج الوصفي الذي يعتمد على تحديد المستوى اللغوي المدروس في زمان ومكان محددين، فضلاً عن المنهجين التجريبي التطبيقي والإحصائي، فالتجريبي التطبيقي يعتمد على علم الأصوات الفيزيائي في التحليل الأوكوستيكي للنصوص اللغوية باستعمال البرامج الحاسوبية وتحليل نتائجها على وفق ما هو متعارف عليه في هذا العلم. أما المنهج الإحصائي، فهو مناسب جدًا للمستوى الصوتي ومتعلق بشكل أساس بتحديد النسب العامة ونسب الأنزياحات بإضفاء روح على التحليل التجريبي الحاسوبي رابطًا بين الفونيمات في الوحدة اللغوية الواحدة، وبعد ذلك بين الوحدات فيما بينها.

ونحن في طريقنا إلى معرفة ما يتبلور في التحليل من قيم صوتية، لا بد لنا من التسليم بأن عناصر الصوت هذه تدفعنا إلى التعامل مع محورين يتمثل الأول: بالأصوات المكتوبة على صفحة ترى بالعين، والثاني: ما ينتجه المتكلم من أصوات في أثناء تلفظه، فنحن إزاء نوعين هما المواد الصوتية و الكتابية، ويهتم الباحث اللغوي بالبحث عن النظام اللغوي العام الذي تكمن وراءه الظواهر اللغوية المختلفة، في الخطاب، وفي الشعر،... الخ من أنواع النصوص اللغوي.

ثانيًا: مبادئ المعالجة الآلية الحاسوبية في المستوى الصوتي

يسهم البحث العلمي اللغوي في تكنولوجيا المعالجة الآلية في تحقيق المبادئ الآتية:

1- إخضاع مظاهر الكلام لبرامج الحوسبة الصوتية والتحليل الطيفي للكشف عن النسيج اللغوي عبر تحليل طبيعة بناء أصواته فيزيائيًّا ونوعية مقاطع الكلمة وتأثيرها في وضوح الكلمات وانسجامها مع بقية المفردات.

2- التوصل إلى نتائج دقيقة من خلال البرامج الحديثة المتطورة المتصلة بالمجالات الصوتية على مستوى الفونيتك (Phonetic) أو على مستوى الفونولوجي (Phonology)، التي لا يعتورها في غالب الأحيان التخمين والتقريب في تحديد الآلية التي تؤدى بها الوحدات الصوتية؛ فإن البنية الصوتية في مظاهر الكلام لم تتحدد بشكل واضح، وخير دليل على ذلك قلة الدراسات التي تبنت هذا الجانب.

3- وتضيف المعالجة الآلية للمستوى الصوتي دراسة الأصوات بنوعيها: طبيعية ولغوية؛ فالطبيعية تظهر في جانبين: جانب فيزيائي (أكوستيكي)، وآخر نطقي (فسيولوجي)، واللغوية تتعلق بالصوت داخل بنية لغوية معينة حاملة لفكرة أو دلالة بحيث تكون تلك العناصر قابلة للوصف والتحليل في ضوء ما توصلت إليه الحوسبة الصوتية.

4- فسح المجال أمام طلبة اللغة العربية وآدابها للاطلاع على التخصصات الفتية التي توجد بالجامعات في عصر العولمة والتطور التكنولوجيا ومدى استغلال الوسائط في التعليم الحديث.

ثالثًا: الصعوبات التي تواجه الباحث في المعالجة الآلية للأصوات

تُعد المعالجة الآلية في تفسير الاشارة الصوتية، وتحليلها أكوستيكيا من أصعب مهام التحليل لاختلاف مواصفات الصوت من متكلم إلى آخر واختلاف النطق من حيث السرعة وربط الكلمات.

وتتمثل الصعوبات، في:

1- قلة المراجع والمصادر والدراسات التطبيقية التي تدرس العلاقة القائمة بين قدرة النص التأثيرية والبنيات الصوتية، فكثير من البحوث لا تقوم على نظريات علمية تُوضِّح مشكلة البحث ومُتغيراته.

2- جنوح بعض النقاد والدارسين المحدثين إلى استعمال الجانب الصوتي في نقد النصوص ودراستها، فإفتقار بعض الباحثين لآليات ومهارات البحث العلمي الآلي يُخرِج بحوثًا ضعيفةً متسرِّعةً مشكوكًا في نتائجها، فبات لزامًا أن تأخذ مثل هذه الدراسات مكانها في كل دراسة لغوية لتكون النتائج أدق وأكثر موضوعية.

3- عدم وجود إستراتيجية وخريطة بحثية واضحة وقاعدة بيانات لبحوث المعالجة الآلية، لذا لابد من تحديد البرامج الحاسوبية المناسبة لكل بحث، والتدرب عليها، وكيفية توظيفها في خدمة أهداف البحث.

4- تنوع الظواهر الصوتية، يؤدى إلى أنّ لكل ظاهرة طريقتها الخاصة بها التي تختلف عن غيرها من الظواهر الأخرى في جانب التحليل في الدراسة التطبيقية لعلم الأصوات الحديث، فعلى الباحث الوقوف على الخطوات اللازمة لقيام إجراء عملي يحدد البنية من دون أن يعزلها عن سياق النص، ومن ثمَّ فهو إجراء متعدد الأبعاد وإن ارتكز على أرضية البعد الصوتي.

رابعًا: أشهر البرامج الحاسوبية المستعملة في المعالجة الآلية للأصوات

لقد اختيرت البرامج الحاسوبية على اعتبار الدقة والجودة في العمل، وهذه البرامج هي:

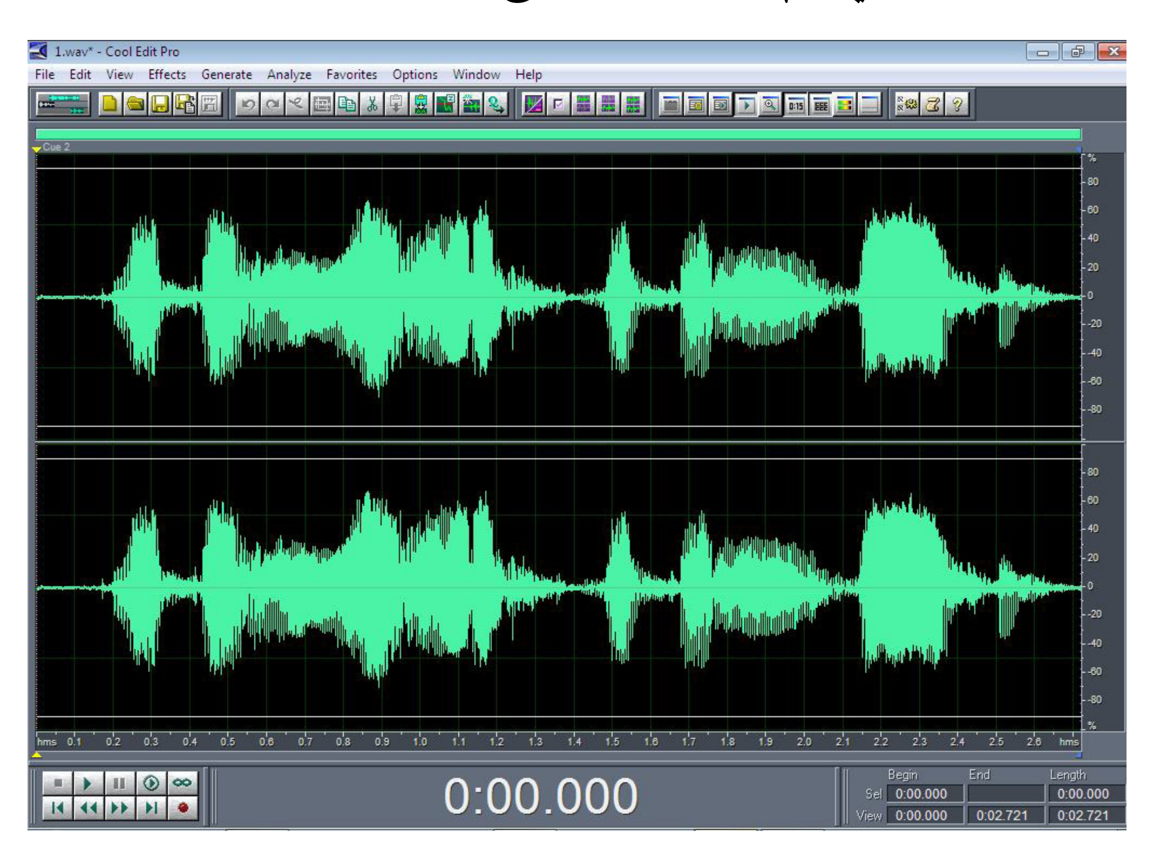
1- برنامج "كول إديت برو" (Cool Edit Pro 2.00).

يستعمل لتجهيز نصوص المادة المراد تحليلها وحفظها في ملفات جديدة ليسهل التعامل معها عند دراستها حاسوبيا؛ إذ يوفر هذا البرنامج إمكانية إنشاء قواعد البيانات الصوتية، بتجهيز الصوت تجهيزًا محكمًا حتى يصبح جاهزًا لأغراض البحث الصوتي، بوساطة التقطيع الصوتي للنصوص في المادة المراد تحليلها وحفظها في ملفات، إذ يتعامل هذا البرنامج مع ملفات الصوت الطبيعي (wave) وملفات الصوت (MP3)، ويتيح امكانات عديدة، منها:

- تسجيل ملفات صوت من الميكروفون أو أي مصدر صوت آخر.
- تشغيل ملف الصوت أو أي جزء منه وتقطيع هذه الملفات وحفظها وإنشاء ملفات جديدة.
- عرض نافذة توضح شكل الموجات (waveform) الخاصة بملف الصوت.
- تسجيل طول زمن المادة الصوتية ويبدأ الوقت من بدء التسجيل أو تشغيل الملف، ويساعد الوقت في أمور كثيرة، ويقاس بالملي ثانية.
- يتوقف التسجيل وتظهر المادة الصوتية على واجهة البرنامج عند الضغط على المسافة من الكيبورد أو التوقف من واجهة البرنامج .
- مدعوم بالأوامر التقليدية لشاشة المفاتيح الخاصة بالويندوز، وكذلك مختصرات الماوس.



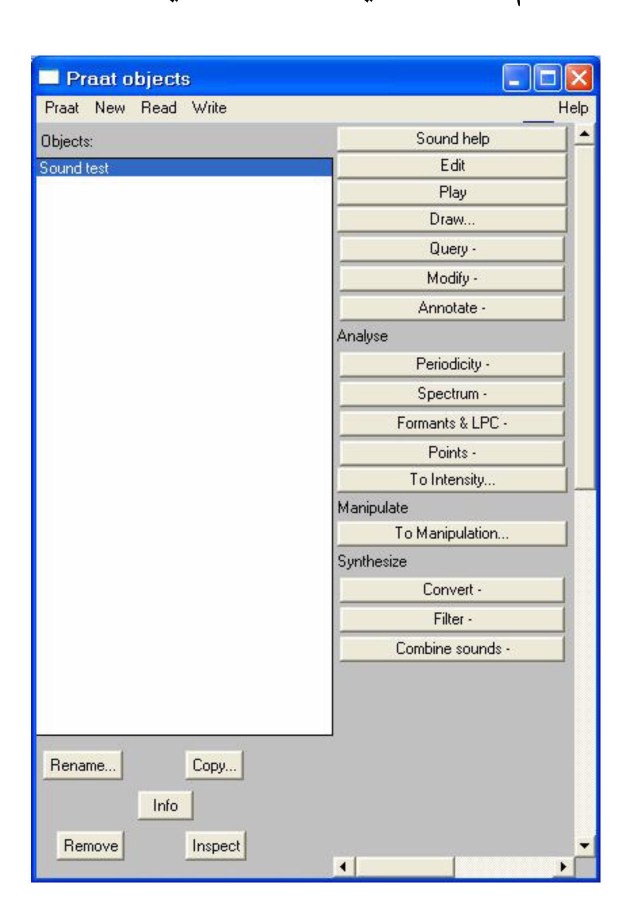
شكل توضيحي رقم (2): واجهة برنامج (Cool Edit Pro 2.00)



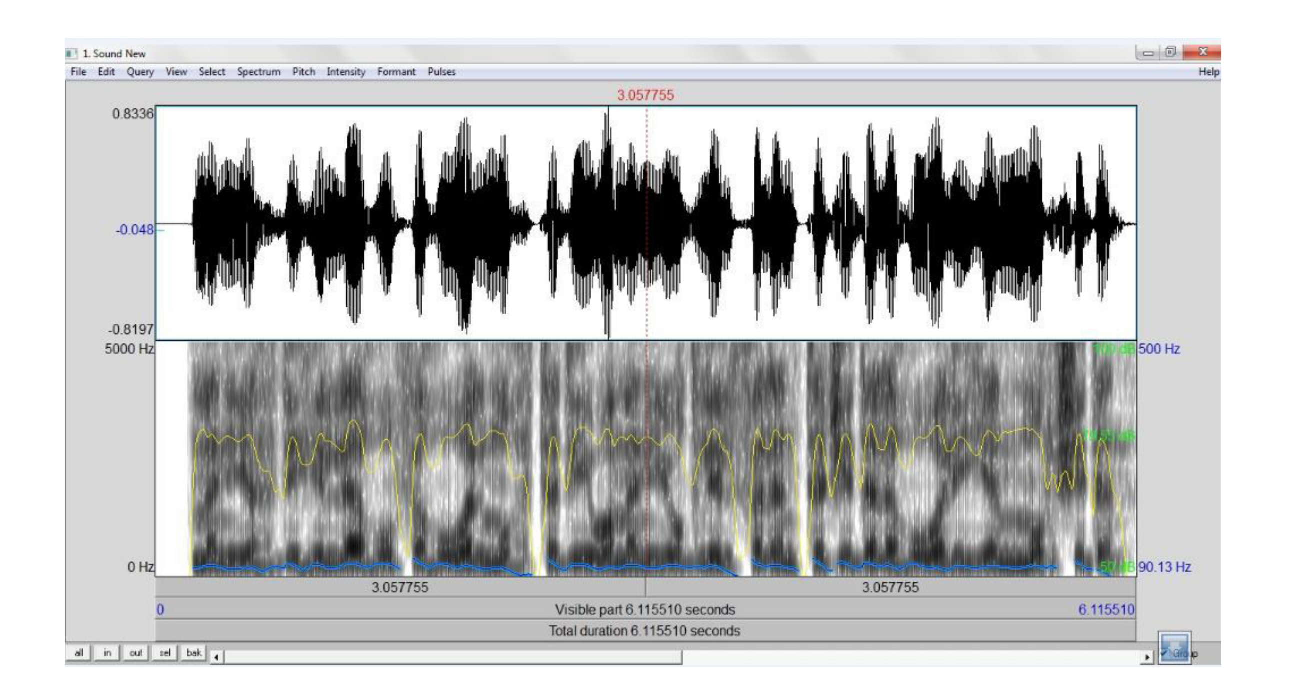
شكل توضيحي رقم (3): طريقة عمل برنامج (Cool Edit Pro 2.00)

2- برنامج برات (praat):

وهو من برامج التحليل الصوتي الأكوستيكي، الذي يتيح إمكانية تحليل منحنى تـردد الأسـاس (Fundamental frequency contour) مـع إدراج الـشكل المـوجي (wave form) لسلـسلة الأصـوات الكلاميـة المحللـة؛ إذ تعـرض الـصورة التحليليـة الناتجـة عنـه معلومات صوتية عن التشكل الذبذبي للمنطوق الصوتي، إلى جانب عرض معلومات رقميـة عن المدى الزمني (Duration) المستغرق للجمل المحللة وقطوعها الصوتية، وعرض البيانـات الخاصة بمنحنى الشدة الصوتية (Intensity) وتحديد قيم القمم النبرية علـى قـوائم البيانـات الداخلية للبرنامج. ومن هـذه البيانـات الأكوسـتيكية الـتي يعنـى بتوافرهـا لدراسـة الظـواهر الفونولوجية، هي: النبر والتنغيم والوقف في النص اللغوي.



شكل توضيحي رقم (4): واجهة برنامج (praat)



شكل توضيحي رقم(5): برنامج (praat)

3- برنامج سناكيت (Snagit. 10) [1].

برنامج مساعد لتصوير التحليل الصوتي من برامج التحليل، وهو يتيح إمكانات عدة، منها:

1- التقاط صورة لشاشة الحاسوب.

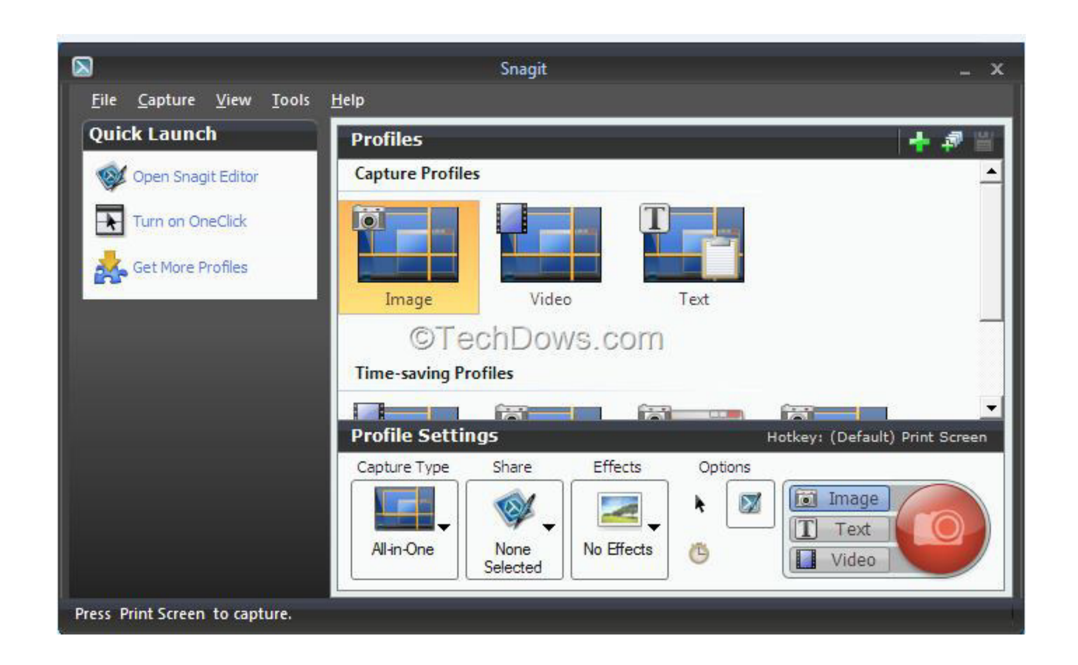
2- الكتابة على الصور.

3- يحتوي على محرر يمكن من خلاله عمل رسومات ووضع أقواس وأسهم وأشكال هندسية ورموز توضيحية.

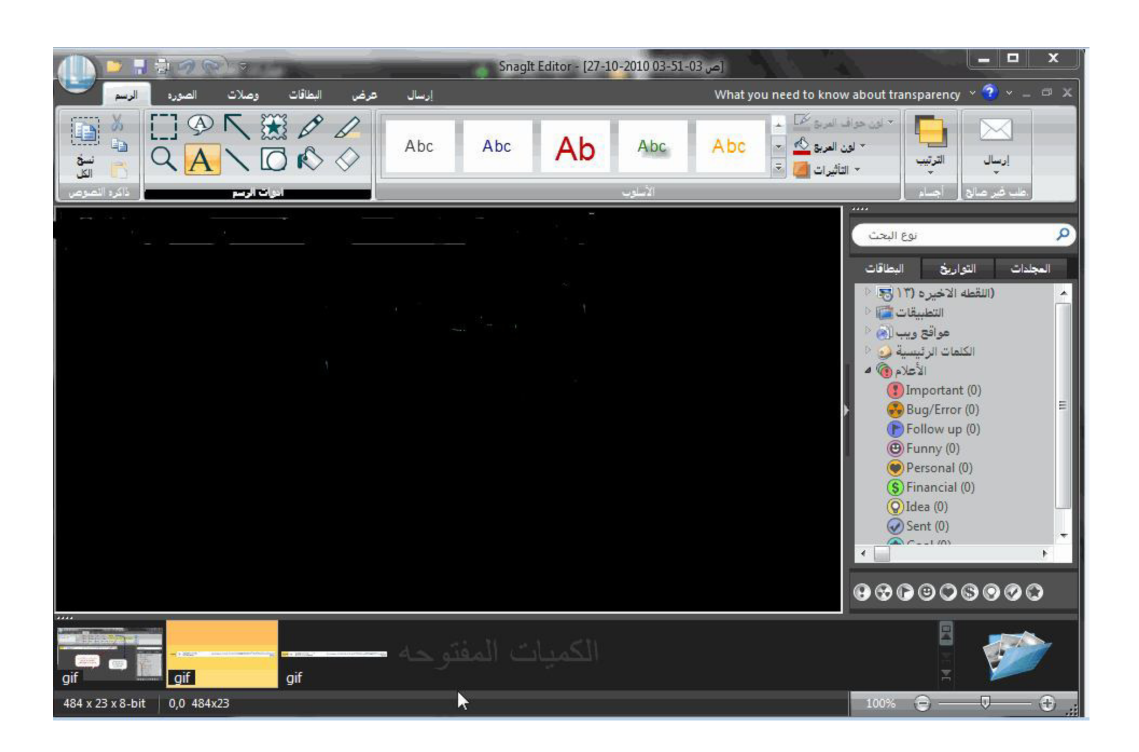
4- إمكانية استدعاء البرنامج- بعد تنصيبه- بالضغط على زر (Print Screen).

---

(1) البرنامج متاح على شبكة الإنترنيت



شكل توضيحي رقم (6): واجهة برنامج (Snagit. 10)



شكل توضيحي رقم (7): برنامج (Snagit)

الخاتمة:

مما لا شك فيه أن المجال اللغوي في مقدمة الميادين الأكثر تأثرًا بتقانة المعلومات، واللغة العربية من بين اللغات المستفيدة، فقد حققت بها قفزة نوعية جعلتها تنخرط في مجال الصناعة اللغوية العالمية.

فاللسانيات الحاسوبية أو ما يدعى بـ تكنولوجيا اللغات أو الهندسة اللسانية، استغل ما توفره التكنولوجيا المُتطورة من أجل بلورة برامج وأنظمة لمعالجة اللغات الطبيعية معالجة آلية، فهو فرع تطبيقي حديث، عرض لآخر النظريات والتطبيقات الحاسوبية؛ بحيث يلتقي فيه الجانب النظري اللساني بكل خلفياته المعرفية والمنهجية مع الجانب التكنولوجي المعلوماتي بكل تطوراته فلا يستطيع أهل اللغة التفرد بتأسيسه ولا أهل الحاسوب كذلك.

سبل تطوير اللغة العربية من خلال اللسانيات الحاسوبية
– المحلل الصرفي أنموذجا.

الدكتورة: بختة تاحي.
جامعة حسيبة بن بوعلي – الشلف – الجزائر

الملخص:

تعتبر اللسانيات الحاسوبية مجالا بينيا حديثا، وهي فرع من فروع اللسانيات التطبيقية أو لنقل فرعا من فروع علم اللغة التطبيقي أو علم اللغة الحاسوبي، حيث كانت بمثابة العامل الأساسي لتطوير وحوسبة اللغة العربية من مختلف مستوياتها، وذلك بالاعتماد على فئتين مهمتين: ألا وهما اللسانيين واللغويين المتخصصين ولا ننسى كذلك المبرمجين الحاسوبيين حيث يعتبر هؤلاء العنصر الأساسي لحوسبة اللغة العربية، عبر مستوياتها التركيبية والدلالية والصرفية والنحوية وغير هذه المستويات، وإنشاء محللات لغوية لهذه المستويات، لدراستها وحوسبتها مما يجعلها قابلة للتداول وسريعة للوصل إلى متناول الباحث أو القارئ.

من هنا يمكننا طرح الإشكاليات التالية: ما لمقصود باللسانيات الحاسوبية؟ وفيما يكمن دورها في معالجة وحوسبة اللغة العربية؟ وماهي أبرز السبل التي اتبعتها لحوسبة هذه الأخيرة؟.

الكلمات المفتاحية:

اللسانيات الحاسوبية- علم اللغة الحاسوبي- اللغويين- اللسانيين- المبرمجين الحاسوبيين- المحللات اللغوية.

Summary :

Computational linguistics is an interdisciplinary field, a branch of applied linguistics or a branch of applied linguistics or computational linguistics. We also do not forget computer programmers, as these are the basic element for the computerization of the Arabic language, through its syntactic, semantic, morphological, grammatical and other levels, and the establishment of language analyzers for these levels, to study and computerization, which makes it negotiable and fast To reach the reach of the researcher or reader.

From here we can raise the following problems: What is meant by computational linguistics? What is its role in the processing and computerization of the Arabic language? What are the most prominent ways you have used to computerize the latter ?.

key words:

Computational linguistics - computational linguistics - linguists - linguistics - computer programmers - language analyzers.

نشأة اللسانيات الحاسوبية :

لقد كانت اللسانيات الحاسوبية في بدايتها تعتمد التحليل الإحصائي للمفردات اللغوية في كتاب معين أو لإعداد فهارس أبجدية لتلك المفردات وتحديد تواترها في مؤلفاته، ثم خطت خطوة مهمة جدا في اختصار الزمن عندما أعدت معاجم إلكترونية أحادية اللغة أو ثنائية أو متعددة اللغات.

وهكذا تم خلال الستينات من القرن 20 وما تلاها حتى الآن إنتاج برامج وأنظمة للترجمة الآلية وغيرها من التطبيقات اللسانية الحاسوبية، كان بعضها تجاريا بحتا، تغذى به الحواسب لتترجم جملا مكتوبة أو منطوقة ومصطلحات كاملة في مجالات محددة أهمها

السياحية والتجارية والمرتبطة بالخدمات كمصطلحات التحية والاستفسار عن الأسعار والأماكن والزمن وغيرها، لكن كغيرها من التخصصات الحديثة والأجنبية المنشأ[1]

يعد حقل اللسانيات الحاسوبية من أوسع الحقول المتصلة بالذكاء الاصطناعي ويسمى الجانب التطبيقي (Artificial intelligence): ويسمى الجانب التطبيقي لهذا الحقل بمعالجة اللغة الطبيعية (Naturel longuage Processing):ولقد جاءت الدراسات في معالجة اللغة العربية متأخرة إذا ما قورنت بالدراسات للغات أخرى وخاصة اللغة الانجليزية ، فاتسمت الأبحاث والدراسات في معالجة اللغة العربية بعدم التنسيق المؤسسي ، فبعضا أخذت الموضوع بالعموم وبعضها درس جزئية محددة ولم تتفق هذه الدراسات السابقة على نموذج محدد لمعالجة اللغة العربية خاصة من الناحية النحوية والمعالجة الآلية للغات تتم ضمن مستويات تحليلية عدة منها : التحليل المعجمي (Lexical Analysis): والتحليل الصرفي (Morphological Analyisis): والتحليل النحوي: (Syntax Analysis) والتحليل الدلالي (Semantic Analysis): قد تم في هذه الاطروحة دراسة هذه المستويات التحليلية الأربعة مطبقة على اللغة العربية التي تم بحثها إلى خمسة:

*الجمل الاسمية الصغرى * الجمل الاسمية الكبرى * وجملة كان وأخواتها * وجملة إن وأخواتها * وجملة الخالفة وقد اقترح الباحث قواعد تركيبية للجمل الاسمية المقترحة في اللغة العربية وعلى أساسها يكون عمل المحلل النحوي (المُعْرِبْ) في معرفة مدى مقبولية الجملة موقع كل مركب فيها من الاعراب.

واستنتج الباحث أن تركيبية الجملة الاسمية في اللغة العربية يمكن الوصول إلى رموزها الصرفية بتحليل تركيبها ضمن مستويات سبعة من ناحية العمق ومن الناحية التطبيقية تم تصميم وبرمجة محلل نحوي للجمل الاسمية غير المشكولة في اللغة العربية باسم (صفوان) ويقوم البرنامج (صفوان) بإجراء التحليل المعجمي والصرفي والنحوي للجمل

---

(1) رضا احمد بابا حوحو: اللسانيات الحاسوبية مشكل المصطلح والترجمة، مخبر المعالجة الآلية للغة العربية، جامعة تلمسان، الجزائر، ص 01.

الاسمية المدخلة اليه ، كما يعطي مجالا لحفظ الجمل والنتائج وفتحهما كملفات نصية ويتمكن البرنامج (صفوان) من اعطاء الاعراب لأجزاء الجملة الاسمية المدخلة، وذلك بالاعتماد على القواعد التركيبية المقترحة لذلك وقد تمت تجربة البرنامج على أكثر من مئة جملة مختلفة التركيب، ونجح في إعراب أكثر من 80 بالمئة منها بشكل صحيح[1].

فقد تمكن باحثون في لغات غربية كثيرة من وضع برامج حاسوبية لغوية طبقوا فيها خوارزميات صورية، وقد استطاعت الآلة أن تتعرف على تلك الخوارزميات وأن تستجيب لها، فظهرت الكثير من البرامج الحاسوبية التي جعلت الحوار بين الإنسان والآلة ممكنا وبلغة طبيعية نذكر منها برامج الترجمة الآلية والمدقق الإملائي والنحوي والقارئ الآلي للحروف العربية المطبوعة أو المكتوبة باليد، تتميز اللغة العربية بجملة من الخصائص تجعلها من أكثر اللغات الطبيعية قابلة للمعالجة الآلية بالحاسوب، فالاطراد في الضوابط والقواعد نجده واضحا في الصرف والنحو والمعجم والأصوات، خلافا لما هو عليه الحال في كثير من اللغات الأخرى، كما أنها تقوم على مكونين رياضيين هما: الجذر والوزن[2].

فروع اللسانيات الحاسوبية:

تشمل اللسانيات النظرية فروعا مختلفة تتناول مستويات متباينة (وقد تكون متداخلة) من التحليل اللغوي، وأهم هذه الفروع.

- علم الأصوات
- علم التصريف
- علم النحو.
- علم الدلالة.

---

(1) نهاد الموسى: العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية ، المؤسسة العربية للدراسات والنشر، بيروت، 2000م ، ص: 47.

(2) عبد العزيز الميهوبي: اللسانيات العربية والحاسوب(معالجة الصرف آليا)، 23 أفريل 2015م، ص 02.

- علم التخاطب[1].

حيث تطمح اللسانيات الحاسوبية – بوساطة توصيف قواعد العربية لأغراض البرمجة الحاسوبية – أن يصبح الحاسوب (محللا لغويا)، قادرا على معالجة اللغة العربية تحليلا وتوليدا، أما أن يكون الحاسوب وسيلة تعليم اللغة العربية أو وعاء تدون فيه اللغة العربية أو مادة درس وسيطة العربية، فإن ذلك يصب في إهتمام مجالات معرفية أخرى، إن الباحث المتخصص في اللسانيات الحاسوبية لا يتحتم عليه أن يكون مهندس حاسوب، يجيد صيانة أجزائه ولا مبرمجا للحاسوب، يتقن لغة البرمجة، ولا مستخدما بارعا لبرامج الحاسوب، يحيط بتفاصيلها ودقائق امكاناتها، إن المتخصص في اللسانيات الحاسوبية لغوي بالدرجة الأولى، يعي النظام اللغوي للعربية ويجيد التعامل مع أمهات الكتب إجادة يعين بها المبرمج على تمثيل اللغة تمثيلا شكليا دقيقا، من شأنه أن يزود الحاسوب بعدة الذكاء الاصطناعي، ويعوضه ما يفتقر إليه من الحدس البشري[2].

فالواقع أن جهود العلماء العرب المعاصرين والمؤسسات العلمية في هذا المجال يمكن نظم عقدها في أربعة صور: الأولى تتمثل في مؤلفات خصصت للعربية والحاسوب، أو الحاسوب والعربية، وجاءت الثانية على هيئة مقالات وبحوث نشرت في المجلات والدوريات العلمية أو ضمن أعمال المؤتمرات، ووقائع الندوات والملتقيات العلمية، أما الثالثة فكانت خاصة بالبرامج والنظم التي وضعت لحوسبة اللغة العربية أو لعوربة الحاسوب، سواء ما كان منها فرديا محض، أو نتاجا مشتركا، أو عملا تجاريا عاما، وأما الصورة الرابعة فتمثلت في إنشاء بعض الكليات الجامعية قسما خاصا لعلم اللغة الحاسوبي كما هي الحال في جامعة الأمير سلطان الأهلية بالرياض (المملكة العربية السعودية).

حيث يُعد كتاب «نبيل علي» (اللغة العربية والحاسوب) أهم الكتب في هذا المجال، وكان أول صدور له عام 1988م، ويمثل هذا الكتاب حجر زاوية في مسيرة البحث اللغوي

---

(1) محمد محمد يونس علي: مدخل إلى اللسانيات، دار الكتاب الجديد المتحدة، ليبيا ، طبعة 01، 2004م، ص 07

(2) وجدان محمد صالح كناني: اللسانيات الحاسوبية العربية – الإطار والمنهج، المعهد الاسلامي للبحوث والتدريب، دت، ص 09.

العربي في اللسانيات الحاسوبية، بل إنه كما وصفه الدكتور نهاد الموسى بحق خطوة واسعة واثقة تنتظم مشروعا لتأسيس اللسانيات الحاسوبية في العربية على أساس نظري وتطبيقي في آن واحد معا.

وبعد نشرها هذا الكتاب بسنوات ثمانية أي عام 1996م، صدر كتاب الدكتور «عبد ذياب العجيلي» (الحاسوب واللغة العربية)، وهو كما يقول «نهاد الموسى»: خطوة جزئية ايجابية نحو معالجة مسائل مفتوحة من العربية بلغة برولوج Prolog وهو يمثل جهدا جيدا في هذا الاتجاه البيني (اللسانيات العربية الحاسوبية) [1].

فقد سعى أولاءك في مؤلفاتهم إلى ابراز مدى أهمية التكنولوجيا الحديثة (الحاسوب) في خدمة اللغة العربية من مختلف المجالات، ومختلف النواحي.

فإن علم اللغة التطبيقي ليس تطبيقيا، لعلم اللغة وليس له نظرية في ذاته وإنما هو ميدان تلتقي فيه علوم مختلفة حيث تتصدى لمعالجة اللغة الانسانية، أو هو علم ذو أنظمة علمية متعددة يستثمر نتائجها في تحديد المشكلات اللغوية وفي وضع الحلول لها، وإذا كان علم اللغة لا يمثل العنصر الوحيد، في ميداننا، لأنه يستقي من علوم أخرى فلا شك أنه يمثل أهم عنصر في [2].

هذا ما أدى به إلى بناء وتوفير تطبيقات، من شأنها أن تساعد في تطوير علم اللغة الحاسوبي، وذلك من خلال إنشاء معدات آلية لمعالجة لهذه اللغة.

- **كيفية بناء تطبيقات علم اللغة الحاسوبي وأنظمة معالجة اللغة الطبيعية:**

ذلك من خلال:

بناء نظام لمعالجة اللغة الطبيعية فذلك يعتبر مهمة صعبة ومعقدة، في الوقت الحالي على أي حال يتمثل التحدي الرئيسي الذي يواجه تصميم هذه الأنظمة في دمج العديد من الأنواع المختلفة للمعرفة – المعرفة النحوية، المعرفة الدلالية، معرفة الكلام (النص)

---

(1) ينظر:عبد الرحمن بن حسن العارف: توظيف اللسانيات الحاسوبية في خدمة اللغة الدراسات اللغوية العربية جهود ونتائج، جامعة أم القرى، دت، ص 03.

(2) بتصرف: عبده الراجحي: علم اللغة التطبيقي وتعليم العربية، دار المعرفة الجامعية، الاسكندرية، دط، 1995م، ص 13.

واستخدامها عمليا في معالجة اللغة في هذا المجال، كما هو الحال عند بناء أي نظام حاسوبي، هنالك بعض التقنيات العامة التي يمكن أن نستعملها لتسهيل عملنا، إحدى هذه التقنيات هي النمذجة (Modularity) وهي تقسيم نظامنا إلى مكونات مستقلة نسبيا، تقسيم المشكلة يسمح لنا بمهاجمة المشاكل الفرعية الأصغر (Sub problems) بشكل مستقل، وبذلك لا تغمرنا المهمة التي أمامنا تماما، إذ ما تم تصميم المكونات بعناية، فسوف نجد أن ذلك التقسيم يقلل ليس فقط من حجم الأقسام المفردة بل من حجم النظام الحاسوبي بأكمله[1].

- **أجزاء أنظمة معالجة اللغة الطبيعية:**

أ- **التحليل والتوليد**: تتطلب أكثر أنظمة معالجة اللغة الطبيعية توفر القدرة على تحليل وتوليد اللغة الطبيعية، يعتبر التحليل المكون الأكثر حسما في الأنظمة العملية التطبيقية لكي يكون نظام اللغة الطبيعية ناجحا، يجب أن يكون قادرا على معرفة وتحليل الكثير من العبارات والصيغ التي تعبر عن نفس الأمر أو المعلومة، من ناحية أخرى يكفي هذا النظام أن يكون قادرا على توليد صيغة واحدة من هذه الصيغ.

ب- **تحليل الجمل وتحليل النص**: بشكل عام ينظر إلى تحليل الجمل على أنه معالجة لجمل فردية وليس تحديدا لبنية الكلام والنص ككل، أي تحليل لبنية الكلام يتطلب تحليلا لمعنى الجمل المفردة في الكلام في كثير من التطبيقات البسيطة لا يكون التحليل الكامل للكلام (النص) ضروريا حيث يمكن أن نفهم الجملة حتى ولو كانت الجملة تشير إلى جمل مذكورة سابقا في سياق الكلام، التحليل الكامل لسياق الكلام يعتبر من أصعب المهام التي تواجه علم اللغة الحاسوبي.

ت- **التحليل النحوي والتحليل الدلالي**: الهدف العام من تحليل الجمل هو تحديد معنى الجملة يتطلب هذا الأمر ترجمة الجمل المدخلة إلى الحاسوب بلغة طبيعية إلى جمل تحمل دلالة بسيطة باستخدام ما يعرف بالمنطق الرمزي، Formal Semantics كذلك يمكن ترجمة الجمل المدخلة إلى لغة مصطنعة خاصة يمكن أن يفهمها الحاسوب مثل لغة

---

[1] ينظر: صلاح الناجم، علم اللغة الحاسوبي، دت، مقال موجود على الموقع: www.alnajem.com، ص 03.

SQL الخاصة بأنظمة إدارة قواعد البيانات في أكثر الأنظمة، أول خطوة لهذه الترجمة الخاصة بالدلالة هي التحليل النحوي، أي عملية تحديد بنية الجملة، وهذه الخطوة هي أول خطوة قام بتنفيذها علماء اللغة الحاسوبيين[1].

إرتبط علم اللغة الحاسوبي أيما ارتباط بالذكاء الاصطناعي، حيث أصبح الحاسوب يحاكي عقل الانسان، من خلال الحوار الشفوي، وسنتبين ذلك من خلال بعض النماذج التي اعتمدت في الذكاء الاصطناعي.

لعبت شبكات الانتقالات دورا هاما في تطوير نماذج الحوار الشفوي بين الانسان والآلة وبالإضافة إلى نماذج Programmar، و Speechis، و Myrtille، التي كانت رائدة في تحليل الخطاب المنطوق وإنتاجه في نموذجين هامين ألا وهما:

- ✓ **نموذج: Lunar**

يعتبر جهاز Lunar (وودس Woods) أول جهاز يعتمد على سؤال وجواب انطلاقا من البحث في قاعدة للمعطيات الخاصة بالمعارف حول الأحجار المستقدمة من القمر ينتج التحليل تمثيلا داخليا على شكل مشجر تركيبي عميق يرتكز على تحليل لساني مؤسس على الأنحاء التحويلية، ولمعالجة الروابط اللغوية، يستعمل ميكانيزما خاصا ينبني على أساليب كشفية خاصة لتفادي اختيار كل التركيبات التي تفرضها عملية البحث، ولا يتدخل المكون الدلالي إلا بعد التحليل التركيبي الذي انجزته شبكة من الانتقالات المعززة، وهو ما يفيد – بالخصوص- في اقصاء التحليلات التركيبية الخالية من المعنى، وفي نموذج أحدث حاول «وودس» Woods تكثيف التفاعل بين التركيب والدلالة مع الحفاظ على استقلاليتهما مبررا ذلك بسببين اثنين:

أولا: إمكانية الحصول بهذه الكيفية على قالب تركيبي مستقل عن أي مجال للتطبيق.

---

[1] ينظر: نفسه، ص 04.

ثانيا: القدرة على تفكيك مشكل التعلم حسب مهمات خاصة، وهو ما يمكن أن ييسر المعالجة الأوتوماتية[1].

✓ نموذج GUS

يتمثل الهدف من نموذج GUS بوبرو Bobrow وآخرون: في تدبير حوارات واقعية بين الإنسان والآلة عندما تكون موجهة للقيام بمهمة محددة مثلا: حوار بين عامل في شركة للأسفار وزبون يرغب في حجز تذكرة طائرة.

الفكرة المركزية لبرامج معالجة طلبات الزبائن هي القابلية، حيث تطابق قوالب الصيغ الصرفية والتركيبية، والدلالية...، التي تتحقق في استقلال عن بعضها البعض، وهذا لا يعني أنها تنطلق بشكل متتابع، بل تتم مراقبتها بواسطة مذكرة تخزن فيها مختلف المهمات الواجب تنفيذها[2].

من بين أهم الإجراءات التي يشغلها هذا النموذج ما يلي:

- التحليل المورفولوجي الذي يعالج سلسلة الحروف في المدخل، وينتج مجموعة الإمكانات لكل كلمة، منظمة على شكل جدول.
- التحليل التركيبي الذي يوظف مفهوم شبكات الانتقالات المتكررة والجدول الذي ينتج عن التحليل الموفولوجي.
- التحليل الدلالي، حيث يولد النحو بنية من الأدوار تحول إلى بنية من الحالات.
- الاستنتاج والتفسير الدلالي: حيث تحلل الجملة حسب المهمة، وتركب عناصر للجواب.
- التوليد الذي يترجم عناصر الجواب إلى اللغة الانجليزية المنطوقة[3].

---

(1) ينظر: بن عيسى زغيوش، مصطفى بوعناني، المعالجة الآلية للغة مسارات تحليل الجمل العربية باعتماد نماذج شبكات الانتقال، مجلة كلية الآداب والعلوم الانسانية، ظهر المهراز، فاس، المغرب، العدد 14، 2006م، ص 66.

(2) ينظر : نفسه، ص 67.

(3) نفسه: ص 68.

اللسانيات وبرمجة اللغة العربية في الحاسوب:

إن التطور السريع والمتنوع الذي تعرفه المعرفة في جميع ميادينها المختلفة يستوجب ضبطها واستعمالها والاستفادة منها عن طريق أجهزة ووسائل عصرية متقدمة تقنية وعلمية مثل الحاسوب الذي يعتبر ذاكرة العصر التي تختزن ما عجزت عنه ذاكرة الإنسان، وإذا كانت دراسة اللغة تستوجب استخدام منهج لساني معين، فإن تخزينه في ذاكرة الإنسان يبدو أمرا صعبا مما يفرض الاستعانة بالحاسوب لمزيد من سرعة العمل العلمي وتحقيق المنهجية والموضوعية وذلك من خلال مستويات (الدلالي، النحوي، الصرفي، المعجمي...) [1].

فقد شملت هذه المعالجة أيضا المستوى الصرفي، في تحليله من جميع الجوانب سواء داخل السياق أو خارجه.

حيث يتناول البحث اللغوي في هذا المستوى الكلمة خارج التركيب فيدرس صيغ الكلمات من حيث بناءها والتغيرات التي تطرأ عليها من نقص أو زيادة وأثر ذلك في المعنى والبحث اللغوي الحديث يتعامل مع المسائل الصرف على أساس صوتي بدلا من اعتماد القدماء على الكتابة اعتبرها القدماء كلمة، في حين يتعامل البحث اللغوي الحديث مع الوحدة الصرفية Morpheme.

أيضا من الحقائق التي أنجزها علم اللغة الحديث تحليل الأنماط الصرفية الخاصة بكل لغة أو لهجة، بقصد الوصول إلى أسلوب كل لغة في بناء كلماتها، كما يهتم علم اللغة الحديث بدراسة التغيرات الصرفية التي تطرأ على بناء الكلمة لاعتبارات صوتية[2].

إرتبط هذا الأخير بالتكنولوجيا الحديثة ونقصد بذلك ارتباطه بالحاسوب، الذي أصبح الوسيلة الفضلى في معالجة الصرف آليا دون اهدار الجهد، وذلك من خلال معالجته من جميع النواحي وتقسيم الكلمة وتحليل بنيتيها الشكلية والعميقة.

---

(1) محمد الحناش: استخدام اللغة في تقنية المعلومات، مجلة التواصل اللساني، مجلد 01، منتدى سور الأزبكية، فاس، المغرب، طبعة 01، 1413هـ- 1993م، ص 31.

(2) محمد محمد داود: العربية وعلم اللغة الحديث، دار غريب للطباعة والنشر والتوزيع، القاهرة، دط، 2001م، ص 106.

1- معالجة الصرف في الحاسوب:

ذلك باستخدام تقنية الآلات منتهية الحالات، حيث ظهر تطبيق لوصف البنية الصرفية للكلمات يعرف باسم الصرف الثنائي المستوى: Two level Morphology، ظهر هذا التطبيق للمرة الأولى على يد "كيموكو سكينيمي" الذي استخدمه لتوليد وتحليل الصرف الفكرة الأساسية وراء الصرف ثنائي المستوى هي أن نفكر في الصرف على أساس أنه عملية تحديد ترابط بين التمثيل السطحي والتمثيل العميق للكلمة، يتم ربط احد المستويين، بالآخر باستخدام الآلات منتهية الحالات، الحزمة المعجمية تتكون من مجموعة للوحدات الصرفية (المورفيمات) التي تظهر في المعجم، الحزمة السطحية، تقابل الصيغة (البنية) التي تظهر عند استخدامها للغة فعليا[1].

فقد سعى الحاسوبيين إلى إنشاء برامج ومحللات تخدم اللغة العربية ومن بين هذه المحللات ما نتحدث نحن اليوم عنه هو المحلل الصرفي والذي يخدم بدوره بحثنا هذا.

2- تصميم المحلل الصرفي:

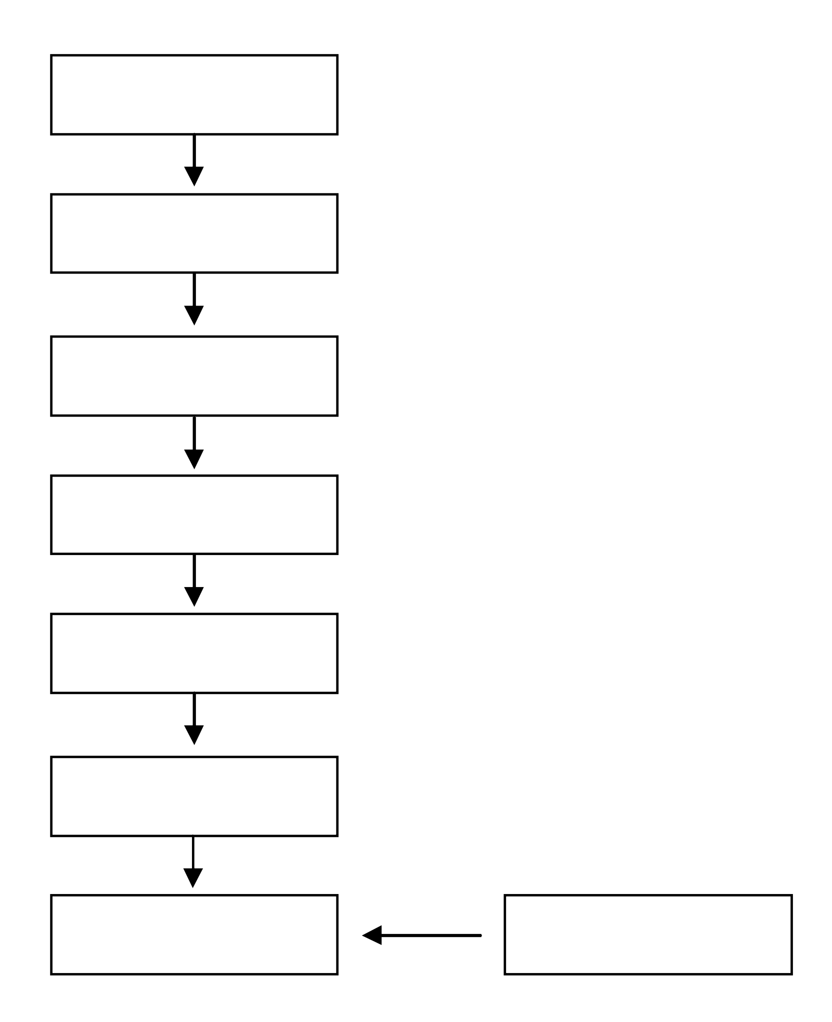
يقوم المحلل الصرفي بتحليل كلمات الجملة المدخلة وتحديد بياناتها الصرفية، وذلك بعد إجراء عدة خطوات.

- تجريد الكلمة من السوابق واللواحق مثل "ال" التعريف والضمائر المتصلة وغير ذلك.
- إيجاد الوزن المقابل لكل كلمة من كلمات الجملة، وذلك من خلال بنى المعطيات.
- إيجاد العلاقات الصرفية النحوية بين مركبات الجملة وذلك طبقا للقواعد النحوية التي تحدد الموقع الصرفي لكل وزن، ويبين المخطط هذه الأوزان من خلال المحلل الصرفي[2]

---

(1) Al- najem , salah, an exploration of comptutional arabic morphology, APhd thesis, exex: university of exex, 1998, p 03.

(2) كريم بوزويع وعبد الفتاح حمداني: وقائع الندوة الثالثة الدولية حول المعالجة الآلية للغة العربية، المدرسة المحمدية للمهندسين، الرباط، المغرب، 2009م، ص 19.



تم تصميم بنى المعطيات للمحلل الصرفي التي تـضمنت أوزان الأسمـاء والأفعـال في اللغة العربية، حيث تم تمثيل المعلومات والقواعد الصرفية لهـذه الأوزان بـشكل معلنـات، وتم وضع التعابير الإعلانية ذات السمات المشتركة ضمن هياكل[1] SLOTS.

تعد اللغة العربية من اللغـات الحاميـة الـسامية في العـالم، حيـث تميـزت بالاشـتقاق وغيرها من المميزات، وبما أنها أصبحت تـشمل جـل العلـوم الانـسانية فقـد ركـزت الجهـود الحديثة إلى ترقية سبل تطويرها وتعليمها، وذلك من خلال الحاسب الآلي، وبما أن علم اللغـة الحاسوبي أو علم اللسانيات الحاسوبية أصبح يعد فرعا مـن فروعهـا، فإنـه لابـد مـن تحقيـق بعض التوصيات في العالم العربي لأجل ترقيتها وجعلها تواكب متطلبات العـصر، ومـن بـين أهم التوصيات:

(1) ينظر : المرجع نفسه، ص 18.

تسخير كامل الامكانيات الماديات والبشرية.

- انشاء فرق بحث لغوية متخصصة في مجال العلوم اللغوية. (صرف، دلالة، معجم، نحو،....إلخ).
- تسخير المعدات الآلية والالكترونية والمعلوماتية لأجل ذلك من أجل تسهيل العملية وتسريعها.
- انشاء مراكز متخصصة في اللغة والحاسوب على حد السواء.

المصادروالمراجع:

1- رضا احمد بابا حوحو: اللسانيات الحاسوبية مشكل المصطلح والترجمة، مخبر المعالجة الآلية للغة العربية، جامعة تلمسان، الجزائر، د.ت

2- نهاد الموسى: العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، المؤسسة العربية للدراسات والنشر، بيروت، 2000م.

3- عبد العزيز الميهوبي: اللسانيات العربية والحاسوب (معالجة الصرف آليا)، 23 أفريل 2015م.

4- محمد محمد يونس علي: مدخل إلى اللسانيات، دار الكتاب الجديد المتحدة، ليبيا ، طبعة 01، 2004م.

5- وجدان محمد صالح كناني: اللسانيات الحاسوبية العربية – الإطار والمنهج، المعهد الاسلامي للبحوث والتدريب، د.ت.

6- عبد الرحمن بن حسن العارف: توظيف اللسانيات الحاسوبية في خدمة اللغة الدراسات اللغوية العربية جهود ونتائج، جامعة أم القرى، دت.

7- عبده الراجحي: علم اللغة التطبيقي وتعليم العربية، دار المعرفة الجامعية، الاسكندرية، دط، 1995م.

8- صلاح الناجم، علم اللغة الحاسوبي، دت، مقال موجود على الموقع: www.alnajem.com.

9- ينظر: بن عيسى زغيوش، مصطفى بوعناني، المعالجة الآلية للغة مسارات تحليل الجمل العربية باعتماد نماذج شبكات الانتقال، مجلة كلية الآداب والعلوم الانسانية، ظهر المهراز، فاس، المغرب، العدد 14، 2006م.

10- محمد الحناش: استخدام اللغة في تقنية المعلومات، مجلة التواصل اللساني، مجلد 01، منتدى سور الأزبكية، فاس، المغرب، طبعة 01، 1413هـ- 1993م.

11- محمد محمد داود: العربية وعلم اللغة الحديث، دار غريب للطباعة والنشر والتوزيع، القاهرة، دط، 2001م.

12- Al- najem , salah, an exploration of comptutional arabic morphology, APhd thesis, exex: university of exex, 1998.

13- كريم بوزوبع وعبد الفتاح حمداني: وقائع الندوة الثالثة الدولية حول المعالجة الآلية للغة العربية، المدرسة المحمدية للمهندسين، الرباط، المغرب، 2009م.

اللسانيات الحاسوبية وفاعليتها في الصناعة المعجميّة الآلية

الدكتورة زينة قرفة

إسلام حب الدين

جامعة محمد البشير الإبراهيمي برج بوعريريج الجزائر

الملخص:

إن هذا البحث يسعى إلى الكشف عن دور اللسانيات الحاسوبية في صناعة المعجم الآلي، إذ أنها الوسيلة الرئيسة في جسر الفجوات المعلوماتية التي تفصلنا عن دول العالم المتقدم وما وصلوا إليه في هذا المجال، إذ وجب علينا استثمار تقنيات هذا العلم وتطويعه لخدمة الدراسات اللغوية العربية في جميع مستوياتها بما في ذلك المستوى المعجمي، ومدى إفادتها منه في معالجة قضاياها المختلفة، كما يبين لنا أيضا كيف أن اللسانيات الحاسوبية ستثري الخزينة اللسانية المصطلحية العربية آليا، إذ يبرز لنا كيفية ربط البحث اللساني العربي المعجمي والحاسوب وفق أحدث التقنيات، وذلك من خلال ما ستقدمه اللسانيات الحاسوبية من خدمة جليلة للمعجم، وتساعدنا في وضع معاجم متطورة نجمع فيه تراثنا اللغوي العربي، ودليل ذلك ما قدمه مشروع الذخيرة اللغوية، والمعجم التاريخي للغة...وغيرهما؛ ومما تقدم يحق لنا أن نتساءل عن علاقة اللغة العربية بالحاسوب؟، وإلى أي مدى تسهم اللسانيات الحاسوبية في تطوير اللغة العربية من حيث مستواها المعجمي؟، وما هو السبيل إلى منهجية فعالة لإعداد معجم لساني حاسوبي عربي؟.

تقديم:

شهدت اللغة العربية في حقبتها التاريخية جهودا ثمينة، شكّلت بذلك مكتبة لغوية تحمل في رفوفها تراثا عريقا لا يضاهيه تراث الأمم الأخرى، فقعدوا لنظرية لغوية في مختلف أنظمتها من صوت وصرف ونحو ومعجم، وهذا الأخير لقي اهتماما بالغا من قبل العلماء، وبذلوا فيه جهودا جبّارة في جمع المادة اللغوية وتدوينها في معاجم حفاظا على لغة القرآن الكريم وهدي النبي صلى الله عليه وسلم والتراث العربي الزاخر، وخوفا من الضياع والاضمحلال، لأن المعجم بؤرة الأنظمة اللغوية الأخرى، فكانت بحق مكتبة غنية باتجاهاتها وأنماطها، أرسى علماؤها قواعد فنية وفق مناهج مختلفة، وتوالت الجهود وتنوعت، وأصبح علما مستقلا بذاته، له حدوده الخاص به، وله منهجه الذي يفصله عن بقية العلوم، وظل هذا العلم رهين التطور إلى أن جاءت فترة تميّزت بنوع من الركود والجمود...، لكن سرعان ما استفاق الذهن العربي في عصر المعلوماتية والرقمنة، وتكاثفت الجهود الفردية والجماعية والمؤسسات والهيئات التي لا تقل أهمية على سابقتها، بل تزيدها تنظيما وترتيبا وسهولة؛ أدت إلى إخراج صورة جديدة للمعجم العربي ماثلة في معجم آلي أو إلكتروني، حتى يسهل على الدارسين والباحثين بمختلف مستوياتهم الحصول على مبتغاهم.

ولأجل ذلك كلّه وجدت أن هذا المجال من المجالات المهمة التي تحتاج إلى بحث ودراسة، إذ كان المعجم الإلكتروني ميدان الدراسة في أطروحة الدكتوراه، فاستقرّ بيَ الحال في هذه الورقة البحثية الموسومة بـ: (اللسانيات الحاسوبية وفاعليتها في الصناعة المعجميّة الآلية)، على جملة من المحاور التي أرى أنها من صميم البحث:

أولا: التطور المعجمي العربي:

كان لظهور الدين الإسلامي الحنيف واهتمام المسلمين من وقت مبكر بجمع القرآن الكريم والحديث النبوي الشريف، وتفسير ما أشكل منهما وشرح غريب ألفاظهما دور أساسي في نشأة الدراسات اللغوية بمختلف أنواعها ومستوياتها بما في ذلك تدوين اللغة.

وكانت نتيجة هذه الحركة اللغوية العظيمة تأليف العديد من الكتب اللغوية الصغيرة وتسمى: (الرسائل اللغوية) أو الكتب المفردة؛ في مختلف الموضوعات ذات الصلة ببيئة

الإنسان العربي الطبيعية وحياته وما يحيط به من مأكل ومشرب وملبس...، وقد كانت هذه الرسائل -كما قسّمها حسين نصار- من حيث الموضوعات تنقسم إلى تسعة أبواب: كتب الغريبين والفقه، وكتب اللغات والعاميّ، والمعرّب وكتب الهمز، وكتب الحيوان، وكتب النوادر، وكتب البلدان والمواضيع، وكتب الإفراد والتثنية والجمع، وكتب الأبنية، وكتب الصفات[1]، وتعدّ الرسائل المندرجة ضمن هذه الأبواب أو المجموعات ما يزيد على تسعمائة رسالة (كتاب) [2].

حيث أن مؤلفي هذه الرسائل لم يتقيدوا بمنهج دقيق في ترتيب الألفاظ التي جمعوها، وإن كانوا قد قسّموها إلى موضعات جزئية تتفرع من الموضوع العام إلى الموضوع الجزئي؛ إذ تعد هذه الرسائل اللغوية الحجر الأساس والخطوة الأولى التي مهّدت الطريق لظهور المعاجم، بل كانت المادة الأساسية التي اعتمد عليها مؤلفو المعاجم في القرن الرابع الهجري وما بعده، ومن هؤلاء الخليل بن أحمد الفراهيدي (ت: 175هـ)، الذي ألّف معجمه المشهور (كتاب العين) أثناء مرحلة تأليف الرسائل اللغوية، إلا أن معجمه قد جاء متميّزا عن تلك الرسائل حجما ووضعا، وكان إنجازه القاموسي إنجازا عظيما وفريدا غير مسبوق له، وقد أعمل الخليل فكره في جمع ألفاظ اللغة بطريقة حاصرة تقوم على الحساب العقلي، حتى لا تشذّ عنه كلمة، وتندّ عنه لفظة، فضلا عن درء مظنة تسرب كلمة أجنبية إلى الألفاظ العربية إلى معجمه، ومن ثم فقد ابتكر نظام التقليبات[3]؛ والمقصود بنظام التقليبات هو تقديم الحرف في الكلمة مرّة وتأخيره أخرى، بحيث يحصل من الثنائي على صورتين ومن الثلاثي على ست صور، ومن الرباعي على أربع وعشرين صورة أو تقليبا، ومن الخماسي على مائة وعشرين صورة حاصلة من تقاليبه، وقد ذكر الخليل أن مبلغ عدد الأبنية في كلام العرب اثنا

---

(1) ينظر: حسين نصار، المعجم العربي: نشأته وتطوره، دار مصر للطباعة، مصر، ط2، 1956م، ج1، من ص: 39 إلى ص 213 للاستزادة.

(2) ينظر: أحمد الشرقاوي إقبال، معجم المعاجم تعريف بنحو ألف ونصف الألف من المعاجم العربية التراثية، دار الغرب الإيلامي، بيروت-لبنان، ط1، 1987م، ص ز وح.

(3) ينظر: صلاح راوي، المدارس المعجمية العربية: نشأتها - تطورها - مناهجها، دار الثقافة العربية، القاهرة-مصر، [د.ط]، 1990م، ص 72.

عشر مليونا وثلاثمائة ألف وخمسة آلاف وأربعمائة واثنا عشر؛ أي ما يزيد على اثني عشر مليون لفظا، على أنَّ هذه الألفاظ جُلُّها مهمل وقليل الاستعمال[1].

حيث تُعد هذه المرحلة مرحلة التأسيس الفعلي للمعجم العربي، وبعدها ألّف العديد من اللغويين العرب والمستعربين منذ كتاب العين للخليل معاجم مشابهة اعتبروها مدارس معجمية، حيث ذهب اللغويون المحدثون إلى تقسيمها إلى مدارس عدّة، ونشير هنا إلى تقسيم حسين نصار الذي قسّمها إلى أربع مدارس:

أولها: مدرسة التقليبات: وتتألف خاصة من (العين) للخليل بن أحمد الفراهيدي (ت:175هـ)، و(البارع) لأبي علي القالي (ت: 356هـ)، و(التهذيب) للأزهري (ت: 370هـ)، و(المحيط الأعظم) للصاحب بن عباد (ت: 385هـ)، و(المحكم) لابن سيده (ت: 458هـ) [2].

ثانيها: مدرسة الترتيب الألفبائي: التي أهملت ترتيب الحروف على أساس المخارج، لكنها تمسكت بنظام الأبنية الخليلية، وتضم معجم (الجمهرة) لابن دريد (ت: 321هـ)، و(مقاييس اللغة) و(المجمل) لابن فارس (ت: 395هـ) [3].

وثالثها: مدرسة القافية: لأنها ترتب المداخل حسب الحرف الأخير، وتضم بدورها كتاب (الصحاح) للجوهري (ت: 393هـ)، و(العباب) للصاغاني (ت: 650هـ)، و(لسان العرب) لابن منظور (ت: 711هـ)، و(القاموس المحيط) للفيروزآبادي (ت: 816هـ)، و(تاج العروس) للزبيدي (ت: 1205هـ)، و(المعيار) للشيرازي[4].

وآخرها: مدرسة الأبجدية العربية: ويمثل هذه المدرسة (أساس البلاغة) للزمخشري (ت: 538هـ)، ومعجم اليسوعيين ومعاجم المجامع اللغوية؛ حيث اعتمدت على ترتيب حروف معجمها تبعا لحرف الكلمة الأول مع طرح نظام الأبنية والمقلوبات[5].

---

(1) ينظر: المرجع السابق، ص 90.

(2) ينظر: حسين نصار، المعجم العربي: نشأته وتطوره، ج1، ص 218 - 392؛ وفيها تفصيل عن هذه المدرسة.

(3) ينظر: المرجع السابق، ج2، ص 404 - 477.

(4) ينظر: المرجع نفسه، ج2، ص 484 - 685.

(5) ينظر: المرجع نفسه، ج2، ص 690 - 741.

وفي خضم هذا التطور الذي شهدته حركة التأليف المعجمي كان للطباعة دور تنويري عظيم؛ إذ أسهمت إسهاما كبيرا في تجديد الثقافة العربية وإحياء تراثها الزاخر عموما ومنه التراث المعجمي خصوصا، وذلك بطبع المعروف منها وما كان مخطوطا لتسهيل التداول بين الناس، ولتسهيل عملية البحث لدى الباحثين والدارسين، وقام بعض اللغويين بإعادة ترتيب بعض المعاجم القديمة على حروف الهجاء حسب أوائل الكلمات قصد التيسير وتسهيل الرجوع إلى المادة المراد البحث عنها وتجشيع الطلاب على الإقبال على استعمالها[1].

وقد كان لحركة الطبع دور مواصلة واستمرارية حركة واسعة في مجال التأليف المعجمي، وتتمثل هذه الاستمرارية في ظهور مدرسة القاموسية اللبنانية في مطلع القرن العشرين (20)، حيث شارك فيها العديد من الباحثين اللبنانيين فأسهموا إلى حدّ كبير في بعث اللغة العربية وإحيائها من جديد[2]، ففي سنة 1869م أخرجت المطبعة العربية قاموسا جديدا في جزئين وضعه بطرس البستني وأسماه (محيط المحيط) ورتّبه على حروف الهجاء بحسب الأوائل، وبعد نظرة ثاقبة في كتابه عمد إلى اختصاره في جزء واحد، أطلق عليه اسم (قطر المحيط) [3].

أما في سنة 1889م ألّف سعيد الخوري الشرتوني قاموسه المشهور (أقرب الموارد في فصيح العربية والشوارد)، وفي سنة 1908م أخرج الأب لويس معلوف قاموسه (المنجد) وقد اختصر فيه محيط المحيط للبستاني الذي هو نفسه تهذيب لـ: (قاموس المحيط)، حيث يعدّ هذا المعجم خير معجم مدرسي للعربية في ترتيبه وإخراجه، ومن أشهر القواميس التي تندرج ضمن حركة القواميس اللبنانية قاموس (البستان) لعبد الله البستاني، و(متن اللغة) لأحمد رضا[4].

---

(1) ينظر: عدنان الخطيب، المعجم العربي بين الماضي والحاضر، مكتبة لبنان ناشرون، بيروت-لبنان، ط1، 1994م، ص 47 - 51.

(2) ينظر: حكمت عشلي، المعجم العربي في لبنان، دار ابن خلدون، لبنان، [د.ط]، 1982م، ص 35 و36.

(3) ينظر: المرجع نفسه، ص 114 - 123.

(4) ينظر: المرجع السابق، ص 130، وينظر كذلك: حسين نصار، المعجم العربي: نشأته وتطوره، ج2، ص 716 – 722.

وبعد هذا كله تتوالى حركة متتالية لمجمع اللغة العربية بالقاهرة سنة 1932م بإسهامها في توسعة دائرة التأليف المعجمي، لهدف الحفاظ على اللغة العربية والمساهمة في إثراء الرصيد اللغوي العربي، وقد حقق المجمع إنجازات مصطلحية وقاموسية عديدة منها: (المعجم الوسيط) الذي أصدره مجموعة باحثين في جزئين سنتي 1960م/ 1961م، ذا نمط حديث؛ لا يقلّ في نظامه عن أحدث المعجمات الأجنبية، فجاء محكمَ الترتيب، واضحَ الأسلوبِ، سهلَ التناولِ، مشتملا على صور لكل ما يحتاج شرحه إلى تصوير، وعلى مصطلحات العلوم والفنون، مع ملحق في آخره يتضمن المشهور من الأعلام والأماكن[1].

والحقيقة أن المعاجم العربية الحديثة كمعجم المحيط وأقرب الموارد والبستان والمنجد والوسيط في نظر الناقدين ما هي إلا صورة للمعاجم القديمة.

أما فيما يخصّ القاموس الذي أصدرته المنظمة العربية للتربية والثقافية والعلوم سنة 1889م، وهو (المعجم العربي الأساسي)؛ إذ تميّز بعناية للألفاظ والمعاني المستحدثة والتعابير الاصطلاحية وألفاظ الحضارة والمصطلحات العلمية[2]، ومن هذه القواميس أيضا (المنجد في اللغة العربية المعاصرة)، الذي صدر سنة 2000م، الذي تميّز بتَفَتُّحِهِ على اللغة العربية الحيّة المعاصرة[3].

ونشير هنا إلى (معجم الطلاب) لمحمود إسماعيل صيني وحيمور حسن يوسف، وهو أول قاموس يعتمد على السياق في تحديد معاني الألفاظ، ويتدرج في عرضه لتلك المعاني من المحسوس إلى المجرد ...، ويهدف إلى خدمة دارسي اللغة العربية من غير العرب[4].

---

(1) ينظر: مجمع اللغة العربية، المعجم الوسيط، القاهرة-مصر، ط3، 1985م، ص 10.

(2) ينظر: تأليف وإعداد جماعة من كبار اللغويين العرب، المعجم العربي الأساسي، المنظمة العربية للتربية والثقافة والعلوم، لاروس-باريس، 1989م، ص 13 - 47.

(3) ينظر: أنطوان نعمة وآخرون، المنجد في اللغة العربية المعاصرة، دار المشرق، بيروت-لبنان، ط2، 2001م، ص 16.

(4) ينظر: محمود إسماعيل صيني، وحيمور حسن يوسف، معجم الطلاب - معجم سياقي للكلمات الشائعة، مكتبة لبنان، بيروت، [د.ط]، 1991م، ص 280.

ومن هذه الإنجازات ما ألّفته اللجنة الدائمة للرصيد اللغوي لدول المغرب العربي معجم (الرصيد اللغوي الوظيفي للمرحلة الأولى من التعليم الابتدائي) سنة 1976م، ويتضمن هذا العمل القاموسي الألفاظ المستعملة المتواترة المكونة للرصيد المعجمي الحيّ[(1)].

ثانيا: مفهوم حوسبة المعجم العربي وواقعه:

1- في تعريف اللسانيات الحاسوبية:

تتعد تعريفات اللسانيات الحاسوبية وتسمياتها، ويصعب إعطاء تعريف شامل ودقيق لها، لكن يمكن أن نجمع بين بعض هذه التعريفات والمفاهيم:

فاللسانيات الحاسوبية (Computational Lunguistics)، وتسمى أيضا بعلوم اللغة الحاسوبية، وهي علوم حديثة تستخدم الحواسيب في تحويل النصوص والمعلومات اللغوية إلى لغات الحاسب الرقمية لتحليلها وترجمتها للغاتٍ أخرى[(2)]؛ إذ تعدّ اللسانيات الحاسوبية من أبرز العلوم اللغوية التي طهرت في العصر الحديث.

يتكون علم (اللسانيات الحاسوبية) من عنصرين أساسيين هما[(3)]:

- أولهما اللسانيات؛ وهو العلم الذي يدرس اللغات الطبيعية الإنسانية في ذاتها ولذاتها سواء أكانت مكتوبة أو منطوقة، أم منطوقة فقط، ويهدف هذا العلم أساسا إلى وصف أبنية هذه اللغات وتفسيرها، واستخراج القواعد العامة المشتركة بينها، والقواعد الخاصة التي تضبط العلاقات بين العناصر المؤلفة لكل لغة على حدة.
- وثانيهما الحاسوبية؛ ويقصد بها توظيف الحاسوب بما يحتويه من إمكانات رياضية وحسابية خارقة، وسعة تخزينية هائلة في خدمة اللغة.

---

(1) ينظر: اللجنة الدائمة للرصيد اللغوي، الرصيد اللغوي الوظيفي للمرحلة الأولى من التعليم الابتدائي، تونس، [د.ط]، 1976م، ص 161.

(2) ينظر: نبيل علي، اللغة العربية والحاسوب، مؤسسة التعريب، الكويت، [د.ط]، 1988م، ص 01.

(3) ينظر: باسم أغبر، علم اللسانيات الحاسوبية: تعريفه ومجالاته، مقال منشور في موقع: مدونات الجزيرة، على الرابط التالي: blogs.aljazeera.net، يوم: 3/ 7/ 2017م، اطلع عليه يوم: 15/ 04/ 2019م، في الساعة: 18:30.

فهي إذن؛ نظام بيني بين اللسانيات وعلم الحاسوب المعني بحوسبة الملكة اللغوية، وباختصار فإنَّ: "اللسانيات الحاسوبية هي العلم الذي يبحث في اللغة البشرية كأداة طيِّعة لمعالجتها في الحاسوب"[1].

ونلاحظ أن تعريف اللسانيات الحاسوبية يختلف من باحث إلى آخر، وذلك عائد إلى الحقل الذي يعمل به اللسانيّ ثم التجربة العلمية التي يخوضها.

وفي الأخير يمكن القول بأن اللسانيات الحاسوبية هي ذلك العلم الذي يحاول ربط اللسانيات بالمعلوماتية قصد معالجة اللغات الطبيعية معالجة آلية، تضاهي معالجة العقل البشري للغة.

2- علاقة اللغة العربية بالحاسوب:

نظرا لما يتميّز به الحاسوب من تقنيات حديثة، وما يقدّمه من خدمات جليلة في شتى مناحي الحياة بما في ذلك اللغة، فمنذ ظهور الحاسوب في أواخر الأربعينيات من القرن التاسع عشر وصلتُه باللغة تتوثق وتتأصل، فقد ذهبت الباحثة زغدودة ذياب إلى الكشف عن العلاقة التبادلية بين اللغة والحاسوب؛ معتبرة أنّ: "اللغة تقع في ذروة الموضوعات التي تهتم بها العلوم الإنسانية، والحاسوب هو ذروة التقنيات الحديثة، فكان من المنطقي أن تلتقي اللغة والحاسوب؛ كون اللغة تجسد النشاط الإنساني الذهني في الوقت الذي يتجه فيه الحاسوب نحو محاكاة بعض وظائف الإنسان وقدراته الذهنية، ونتيجة لهذا فإننا نلاحظ أن تكنولوجيا المعلومات قد فجرت إشكالية اللغة فأقيمت المراكز والبحوث للبحث في علاقة اللغة بالمعلوماتية، واستهوت بدورها العديد من الباحثين والمهندسين واللغويين للوصول إلى درجة محاورة الحاسب أو التخاطب معه، لفهم عمليات إدراك الكلام وإنتاجه، والحصول

---

(1) دائرة الحوار، اللغة العربية والحاسوب، مجلة عالم الفكر، ع: 2، يوليو- أغسطس - سبتمبر 1988م، م: 19، ص 248.

على أكبر قدر من المعلومات والخدمات وتطوير نظم التواصل مع الآلة الذي يساعد عامّة النّاس على التفاهم مع الحاسوب"[1].

ويشير باحث آخر إلى الترابط العضوي بين الحاسوبية واللغة في جوانب عدّة؛ إذ الغاية من حوسبة اللغة العربية هي تقديم توصيف شامل ودقيق للنظام اللغوي للحاسوب تمكّنه من مضاهاة الإنسان في كفايته وأدائه اللغويين، فيصبح قادرا على تركيب اللغة وتحليلها، ويعرف قواعد النظام الكتابي فيكشف الأخطاء الإملائية ويبني الصيّغ الصرفية وينشئ الجمل الصحيحة... [2]، وذلك عبر مستويات عدّة: (مستوى الصرف، مستوى الكلمة، مستوى الجملة، ومستوى الصوت، ومستوى المعجم).

3- مفهوم حوسبة المعجم العربي:

عدّ محمود فهمي حجازي حوسبة المعجم من أهم مجالات علم اللغة الحاسوبي وأكثرها تلبية للمتطلبات العلمية والثقافية في الدول المتقدمة في العالم المعاصر؛ إذ "يقدّم الحاسوب خدمات كبيرة للبحث اللغوي والأدبي من خلال المعاونة في إعداد معجمات المدونات والمقصود بمعجمات المدونات كل الأعمال المعجمية التي تقوم على الإعداد المعجمي لمجموع الكلمات الواردة في نص محدد"[3].

وتذهب الباحثة «راضية بن عريبة» إلى أنّ المعالجة الآلية للمعجم العربي أو حوسبته هي: "اعتماد نظم الحوسبة المتقدمة استنادا إلى خوارزميات برمجية تستثمر المنطق المعجمي

---

(1) ينظر: مروش ذياب زغدودة، علاقة اللغة العربية بالحاسوب، مجلة الآداب والعلوم الإنسانية، جامعة الحاج لخضر باتنة -1-، الجزائر، ع: 3، ديسمبر 2009م، ص 109.

(2) ينظر: وليد إبراهيم الحاج، اللغة العربية ووسائل الاتصال الحديثة، دار البداية ناشرون وموزعون، الأردن، ط1، 2007م، ص 32 و33.

(3) محمود فهمي حجازي، المعجمات العربية وموقعها بين المعجمات العالمية، بحوث ندوة خاصة بمناسبة الانتهاء من تحقيق وطباعة معجم تاج العروس، المجلس الوطني للثقافة والفنون والآداب، الكويت، يومي: 09 و10 فبراير 2002م، ص 18.

العربي في معالجة المفردة العربية عن طريق استخلاص العناصر الأولية لبنية الكلمة، ومباشرة تحديد سماتها المعجمية[1].

ويعتمد في ذلك على معالجة مفردة واحدة حسب اختيار المستخدم، ويباشر بإظهار أهم الجذور المتاحة لتلك المفردة في الموسوعة قيد الدراسة، مع إتاحة الفرصة لانتقاء الجذر المناسب لعملية البحث[2].

ويمكن القول بأن المعجم المحوسب أو المعجم الآلي/ الإلكتروني ما هو إلا معجم تقليدي محوسب وممكنن، يخزن على شكل شرائح إلكترونية أو وسائط ممغنطة كالأقراص الممغنطة أو الضوئية[3].

4- واقع حوسبة المعجم العربي:

إقتحم العرب مجال حوسبة المعاجم العربية القديمة من السبعينيات، وذلك بعقد المؤتمرات والندوات العالمية والعربية مثل الملتقى الرابع للسانيات العربية والإعلامية بتونس، والمؤتمر الثاني حول اللغويات الحسابية العربية بالكويت عام 1989م، كما يعدّ على رأس المشتغلين بحوسبة الدراسات اللغوية عبد الرحمان الحاج صالح الذي قدّم عدّة أعمال رائدة حول العلاج الآلي للنصوص العربية، وبحوث عبد القادر الفاسي الفهري حول حوسبة المعجم العربي، ومحمّد الحنّاش الذي قام بدراسات حول المعجم الإلكتروني للغة العربية، واقترح مشروع نظرية حاسوب لسانية في سبيل بناء معاجم آلية للغة العربية، ومحمّد حشيش الذي قدّم بحوثا حول معالجة اللغة العربية بالحاسوب، ونبيل علي الذي قدّم أعمالا رائدة منها: اللغة العربية والحاسوب، وميكنة المعجم العربي باستخدام المعالج الصرفي الآلي وبحثه:

---

(1) راضية بن عريبة، حوسبة النظام اللغوي العربي، مجلة اللغة العربية، الجزائر، ع:37، ص 25.

(2) ينظر: المرجع نفسه، ص 25.

(3) ينظر: أحمد محمد المعتوق، المعجم الآلي: أهميته وسبل مقاربته بالمعجم الورقي، مجلة الفيصل، ع: 290، 2000م، ص 60.

الجيل الخامس ومعالجة اللغة العربية آليا[1]، وقـد أنجـزت أعمـال في هـذا الميـدان مـن طـرف الجامعات العربية والمنظمات التابعة لها، فكانت إسهاما حقيقيا مـن أجـل إثـراء هـذا الحقـل الخصب.

إنّ استخدام الحاسوب في صـناعة المعـاجم وتـصنيفها يعـدّ ظـاهرة مثاليـة للمعالجـة اللغوية، فقد استطاعت الدراسات اللسانية تحقيق تقدم ملموس ظهـرت نتائجـه علـى شـكل معاجم آلية قابلة للاستعمال العادي من قبل الباحثين والمتعلمين، إذ تـساعد حوسـبة المعجـم العربي على تسهيل معجمية الرصيد اللغوي العربي الثري في حافظات برمجية جاهزة للتـسيير وفق الأغراض المعجمية المنشودة من حيث الإحـصاء والوصـف والتعـدد الـدلالي والتـوزع اللغوي الـصوتي أو الـصرفي أو النحـوي أو البلاغـي أو الاصـطلاحي والمجـالات الإبلاغيـة والاتصالية[2].

يعدّ المعجم الآلي بمختلف مستوياته مصدرا لا غنى عنه لدراسـة الإنتاجيـة الـصرفية للغـة العربيـة وعلاقـة الترابـط الأخـرى للعناصـر المعجميـة الأخـرى، وعليـه يتفـق خـبراء الحواسب على وجود نوعين اثنين من المعاجم الآلية العربية[3]:

1- المعجم الآلي للمفردة في مستواها الإفرادي، حيث يتضمن هذا المستوى تقديم وصـف صوري للمفردات المكوّنة للغة حول الكيفيات التي تُرسم بها الحروف المؤلفة للوحـدة المعجمية، ووصف دقيق حول كيفية نطق الكلمات والحركات التي تحدّد رسم المتواليـة اللغوية ووضع رصيد صوري لجميع المعطيات المورفولوجية التي قد تتجلّى من خلالهـا الوحدة المعجمية، وتقديم كلّ المعلومات المرتبطة بالقيم الصرفية والنحويـة، إضـافة إلى

(1) ينظر: أحمد هاشم أحمد السامرائي، حوسبة المعجم العربي والقضايا المعاصرة، مجلة سر من رأى، ع: 34، م: 9، السنة التاسعة، تموز 2013م، ص 11.

(2) ينظر: عبد الله أبو هيف، مستقبل اللغة العربية: حوسبة المعجم العربي ومشكلاته التقنية أنموذجا، مجلة التراث العربي، دمشق، ع: 93 و94، السنة الرابعة والعشرون، آذار وحزيران 2004م/ المحرم وربيع الثاني 1424هـ، ص 114.

(3) ينظر: محمد الحناش، مشروع نظرية حاسوب –لسانية في بناء معاجم آلية للغة العربية، مجلة التواصل اللساني، ع:2، م: 2، 1990م، ص 46 و47.

المعلومات الدلالية التي تحمل من طرف الجذور والمباني الصرفية التي تتجلّى من خلالها الكلمة.

2- المعجم الآلي للمفردة في مستواها التركيبي: هو تكميلي للمعطيات التي تُسجل للمفردة في مستواها الإفرادي، يتم في هذا المستوى تقديم كل المعطيات اللغوية التي يمكن أن تكتنف المفردة من خلال ما يقتضيه المستوى التركيبي الذي يعطي للمفردة وظيفتها اللغوية المناسبة في سائر السياقات المختلفة.

5- الآفاق الاستشرافية لحوسبة المعجم العربي:

على الرغم من الجهود الرائدة في مجال حوسبة المعجم العربي، إلا أن الأمر ليس بالأمهر السهل؛ إذ تعترض طريقه صعوبات تعود إلى عدّة أسباب أهمها:

- واقع البحث المعجمي قليل ونادر في العالم العربي.
- يتطلب إنجاز الأعمال الضخمة كالمعاجم والموسوعات إلى تخطيط محكم وتنفيذ دقيق.
- فقدان التعاون المشترك بين مختلف الأطراف.
- هذه الأعمال تحتاج إلى إنفاق ضخم.

ونرى أنّه من المفيد أن نورد بعض الحلول التي نراها كفيلة بتطوير حوسبة المعجم العربي، وقد أشار إليها بعض الباحثين في دراسات سابقة، وتتمثل هذه الحلول في[1]:

- أنّ أيّة محاولة لبناء معاجم آلية للغة العربية يجب أن تبدأ من الوصف اللساني لنظام هذه اللغة، فاللسانيات كما يقول محمّد الحنّاش هي وحدها القادرة على إعطاء الوصف الصحيح لنظام اللغة، وذلك لأنّها تتبع العملية الإبداعية اللغوية، فقد أصبح يفرض في اللساني تمكنه من القواعد الأساسية لأنظمة المعلومات المعاصرة، ذلك لأنّها أداة فعّالة تدفع باللساني إلى تكييف اقتراحاته لغاية المعالجة الآلية لنظام اللغة،

[1] ينظر: محمد الحناش، مشروع نظرية حاسوب –لسانية في بناء معاجم آلية للغة العربية، ص 41، وينظر كذلك: عبد الله أبو هيف، مستقبل اللغة العربية: حوسبة المعجم العربي ومشكلاته التقنية أنموذجا، ص 117

هذا كلّه يدعو إلى التعاون بين اللسانيين والحاسوبيين العرب من أجل بناء معاجم آلية للنظام اللغوي العربي.

- مما يساعد على وفرة الخبرات المطلوبة لصناعة المعجم العربي أن تساهم المؤسسات الأكاديمية في ذلك عن طريق إدخال برامج جديدة في أقسام اللغات تتعلق بنظرية المعجم وتطبيقاته العملية، وإنشاء دبلومات دراسية تختص بالعمل المعجمي. ويقتضي هذا العمل تشجيع طلّاب الدراسات العليا في أقسام اللغات على توجيه رسائلهم للماجستير والدكتوراه لدراسة المشكلات المعجمية، وخلق قنوات اتصال بين مراكز البحث والتأليف المعجمي في العالم العربي ومثيلاتها في الدول المتقدمة لاكتساب الخبرة، والتزود بالتقنيات الحديثة للعمل المعجمي.
- تطوير عمل المجامع اللغوية في هذا المجال والشروع في البرمجيات لوضع إطار تقانة المعلومات من منظور اللغة وإقامة النماذج اللغوية، وتحليل فروعها المختلفة في ميادين الصرف الحاسوبي، والنحو الحاسوبي، والدلالة الحاسوبية، والمعجمية الحاسوبية للمواءمة بين المنظومات البرمجية وطبيعة اللغة العربية.

ثالثا: أهمية الحاسوب في صناعة المعاجم الآلية:

تتجلى أهمية الحاسوب في الصناعة المعجمية الآلية، فيما يلي:

1- تكمن أهمية الحاسوب في احتوائه على ذاكرة هائلة تمكّنه من تخزين عدد كبير من المعلومات، وترتيبها وفق نظام خاص يحدده الدّارس، وهذه الأهمية سهّلت العمل المعجمي للدارسين، إذ تخزّن المادة اللغوية وفق نظام خاص، فيعدل فيه ما يريد ويحذف ما يريد منها، ونتيجة لذلك فإن عملية الزيادة على المعجم تكون سهلة وميسّرة للدارس، فضلا عن تسهيل مهمة تنضيده وطبعه كاملا أو مجزأ[1].

2- يوفر لنا الحاسوب أرشيفا كاملا من المعلومات اللغوية المرتبة آليا؛ وهي عبارة عن أقراص مضغوطة أو وسائط ممغنطة، أو شرائح إلكترونية يصطلح عليها بـــ: (- CD

(1) ينظر: محمود فهمي حجازي، المعجمات العربية وموقعها من المعجمات العالمية، ص 18.

(ROM)، أو (Flash Memory)؛ إذ بدورها توفر لنا سعة ذاكرة ضخمة لا يمكن حفظها في مجلدات ورقية، وتوفر لنا السرعة في عملية البحث وعملية الاسترجاع عن المادة المطلوبة في جزء من الثانية، ثمّ يحقّق المنهجية والموضوعية في الأعمال اللغوية، ويدفع الباحث اللساني لأن يكون دقيقا وموضوعيا وسريعا في بحوثه اللغوية[1].

3- استخدام الحاسوب في الصناعة المعجمية يسهم في تطوير الترجمة الآلية وتحسينها، يرتد ذلك إلى شمولية النظام اللغوي الذي يزود به الحاسوب، إذ يستطيع أن يضع أمام المترجم المعنى الدقيق للمصطلح في كلّ فرع فروع المعرفة، فمن المعروف أنّ معنى المصطلح أو مدلوله يتغير طبقا لحقل الاختصاص الذي يُستعمل فيه في حين يعجز معجم واحد عن سرد معاني المصطلحات المتنوعة في مختلف فروع العلوم والتكنولوجيا[2].

4- تحليل العلاقة بين مفردات المعجم وعناصره كالعلاقة بين جذور الكلمات والصيغ الصرفية أو قواعد تكوين الكلمات المطبقة عليها أو العلاقات الموضوعية التي تجمع بين عائلات المفردات (مصطلحات رياضية، مصطلحات طبية...) [3].

رابعا: منهجية إعداد معجم آلي:

يتطلب بناء معجم آلي وفق منهجية فعالة تؤسس لمعجم يسهل استعماله ونشره وتداوله، إلماما كبيرا بجميع مستويات اللغة، والاعتماد على خطوات وآليات وأدوات ورموز مشفرة، ثم قاعدة معطيات دقيقة ومحكّمة ومنظمة تشمل مستويات اللغة، إذ نحاول هنا التعرف على جملة الآليات والخطوات التي يقوم عليها إعداد معجم آلي، وذلك عبر مساءلة مجموعة من الباحثين والدارسين؛ وتتمثل هذه الخطوات فيما يلي:

---

(1) ينظر: أحمد السامرائي، حوسبة المعجم العربي والقضايا المعاصرة، ص05 و06.

(2) ينظر: المرجع نفسه، ص 06.

(3) ينظر: نبيل علي، اللغة العربية والحاسوب، ص 93.

1- المادة اللغوية أو مداخل المعجم:

حيث يتطلب إعداد معجم إلكتروني للغة العربية الاعتماد على مصادر تكون بمثابة قاعدة أساسية لثراء المعجم بمختلف المعلومات المراد البحث عنها، فهذه المصادر تكون بمثابة المدونة الأساسية للمعجم الإلكترووني، ومن جملة هذه المصادر:

- معجمات اللغة العربية القديمة والحديثة.
- الدواوين الشعرية.
- القرآن الكريم والحديث النبوي الشريف.
- ما أقرّته المجامع اللغوية من ألفاظ ومصطلحات وعبارات وأساليب.
- بعض الصحف والمجلات.

وعليه؛ فإن مشكلة الحجم لا تمثل عائقا أمام المعجم الحاسوبي، إذ يمكنه جمع كل ألفاظ اللغة، ويقوم على معالجتها، ويختار مستعمل المعجم المدخل الذي يريد البحث عنه.

2- ترتيب المداخل:

يمثل ترتيب المداخل الخطوة الثانية في إعداد معجم آلي، إذ يسعى هذا الأخير إلى تفعيل خطوة هامة وهي الدمج بين الترتيبين اللذان يقوم عليهما المعجم الورقي (ترتيب المداخل حسب الألفاظ، أو ترتيب المداخل حسب المعاني)؛ إذ يستطيع المعجم الإلكتروني أن يقدّم عملية البحث باستخدام اللفظ للوصول إلى المعنى (معجم ألفاظ)، واستخدام المعنى للوصول إلى اللفظ (معجم معاني)، فيمكن للمستعمل أن يبحث عن كلمة (قرأ) على سبيل التمثيل، كما يمكنه اللجوء إلى البحث عن الألفاظ الدالة على معنى القراءة[1].

---

(1) ينظر: عمر مدكور، المعجم بين الورقية والحاسوبية، مقال منشور في كتاب أعمال المؤتمر الدولي الأول حول: (اللغة العربية وآدابها: نظرة معاصرة)، قسم اللغة العربية، جامعة كيرالا، الهند، 2015م، ص 179.

3- معالجة الضبط والهجاء:

تعدّ هذه الخطوة خطوة أساسية في عملية إعداد معجم آلي، ولا تختلف المعاجم الحاسوبية في طريقة هجاء وضبط مداخلها كثيرا عن المعاجم الورقية، تجنبا للوقوع في التحريف أو الخطأ في الشكل، لكن المعجم الورقي يستطيع أن يقدّم بعدا آخر لمعالجة عملية الضبط والهجاء، وهو الضبط الصّوتيّ للمدخل، وبذلك يكون قد ساعد في المعالجة الدقيقة للمدخل، فالهجاء معالجة كتابية بصرية، أما ضبط المدخل فمعالجة صوتية، تقدّم بصورة صوتية عبر برامج النّطق الآلي للنصوص المكتوبة، أو تسجيلا صوتيًّا حيًّا للمداخل[1].

4- معالجة المعنى:

عادة ما يلجأ المعجمي أثناء معالجته للمعنى في المعاجم الورقية لأكثر من طريقة من طرق الشرح، والتوضيح عن معاني المداخل اللغوية، وأهمّها: التعريف بالشرح، الشرح بالمترادفات، والمضاد، والشرح بتحديد المكونات الدلالية والسياق، والشرح بالرسوم والصور.

ويأتي دور المعجم الحاسوبي لتقديم معاني هذه المداخل بالطرائق نفسها التي يلجأ إليها المعجم الورقي، إلا أنّه يقدّمها بشكل مغاير؛ إذ يمكنه استخدام وسائط متعددة تعمل على تقديم المعنى للمستعملين بصورة أكثر وضوحا، ويضيف المعجم أيقونة تقدّم وسيلة إضافية لشرح المعجم تستخدم الوسائط المتعددة، صوتا وصورة وحركة على شكل: فيلم متحرك، أو صورة متحركة، أو صورة صوتية، وتقديم الأمثلة الظاهرية بصورة حيّة، وليست هذه الوسائط هدفا قائما لذاته، وإنّما هي وسيلة لإظهار المعنى بأقصى درجات الوضوح للمستعملين والباحثين[2].

---

(1) ينظر: المرجع نفسه، ص 180.

(2) ينظر: المرجع السابق، ص 180 – 184.

5- تنظيم قاعدة معطيات المعجم:

بعدما تم جمع المعطيات اللغوية أو ما يعرف بالمدونة اللغوية المستقاة من المصادر السابقة الذكر، توضع في قاعدة معطيات (ACCESS)، وهذه القاعدة مستقلة عن برامج المعجم[1].

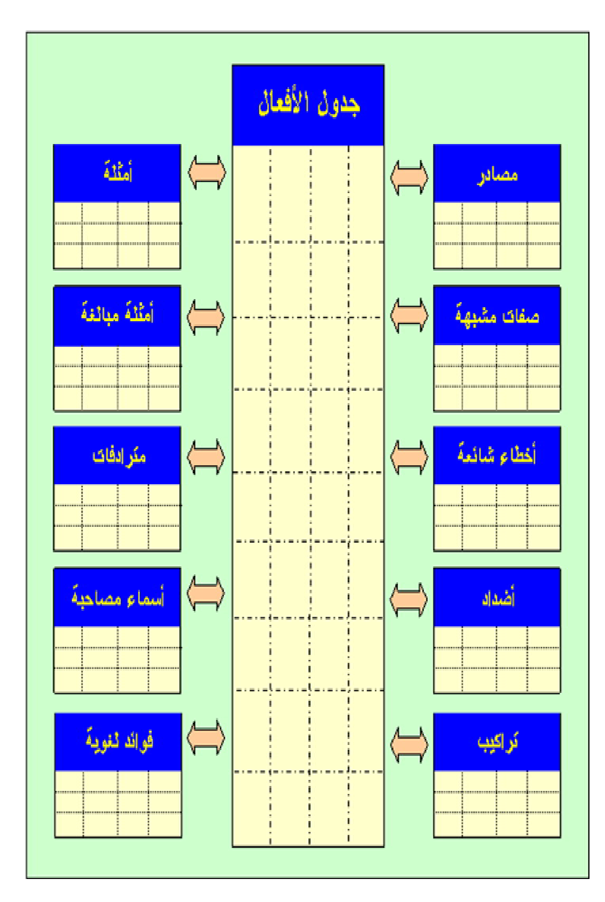
إذ تتكون قاعدة المعطيات من أربعة جداول أساسية هي: (جدول الأفعال، جدول الأسماء، جدول الأدوات، جدول الإحصاء)، كما هو موضح في الشكل التالي[2]:

(1) ينظر: مروان البواب، منهج إعداد معجم العربي الحاسوبي للغة العربية، الاجتماع الثاني لخبراء المعجم الحاسوبي التفاعلي للغة العربية، 1429هـ/ 2008م، ص 03.

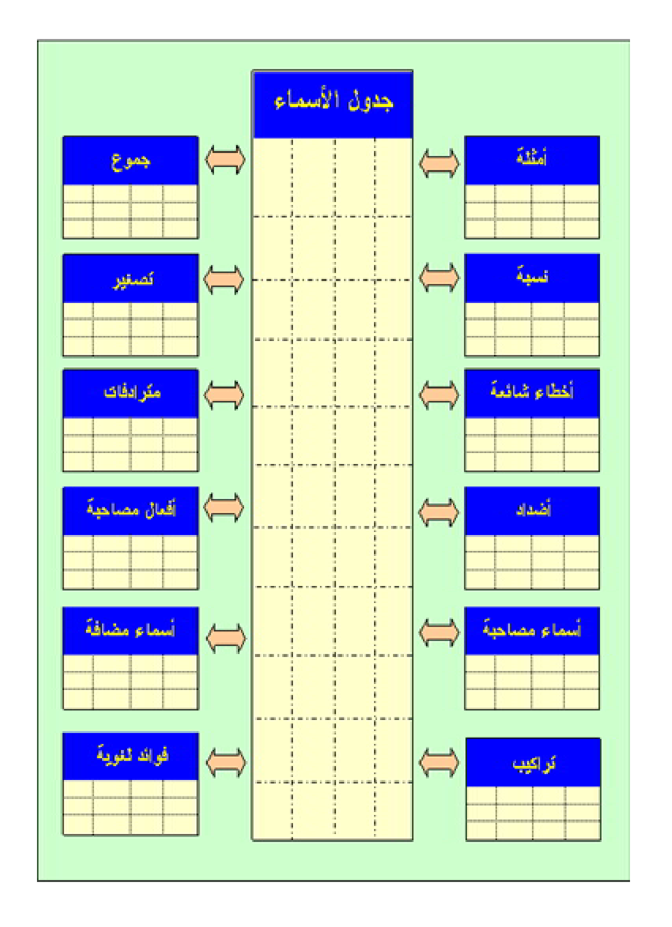
(2) ينظر: مروان البواب، أثر التقانات الحديثة في تجديد المعجم العربي، مقال منشور في موقع: www.alukah.net، يوم: 18/ 03/ 2009م، اطلع عليه يوم: 27/ 04/ 2019م، في الساعة: 14:00.

ويرتبط بجداول هذه القاعدة جداول فرعية، نحاول الإشارة إلى كل واحد منهـا علـى حدة بشكل توضيحي[1]:

أ- جدول الأفعال:



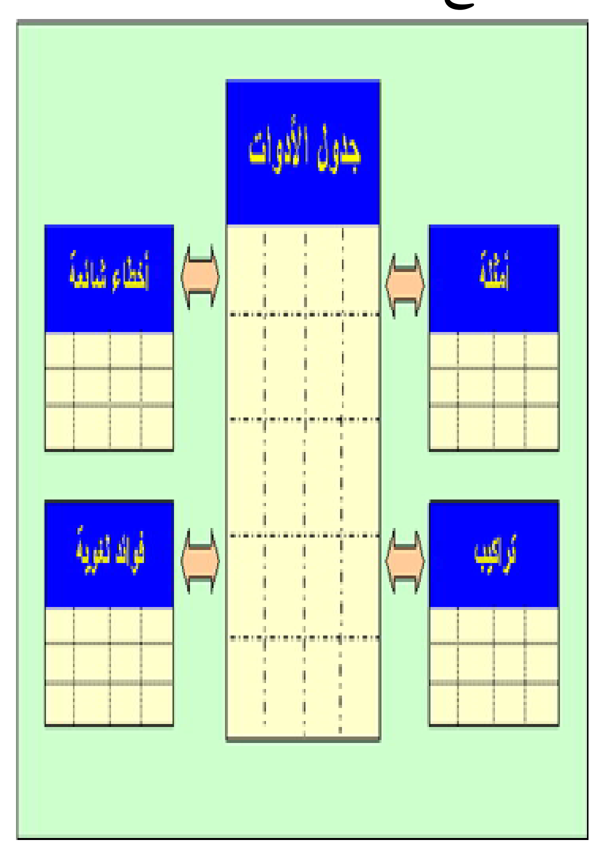
ب- جدول الأسماء:



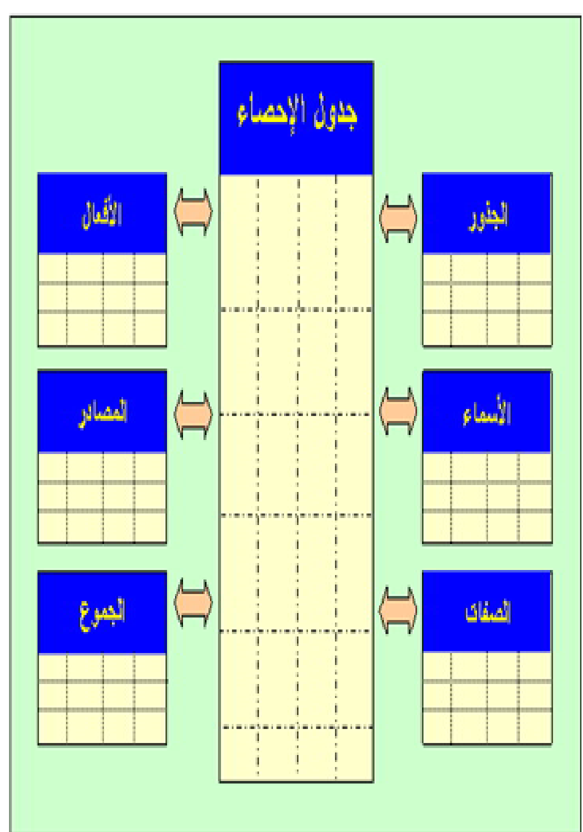
---

(1) المرجع نفسه.

ج: جدول الأدوات:



د. جدول الإحصاء:



حيث تمكننا هذه الجداول من إجراء عمليات بحث مختلفة في المعجم الإلكتروني.

كما أن عملية الحوسبة المعجمية تقوم على منظومتين أساسيتين هما:

- الوحدات الصوتية.
- تأليف الكلام العربي.

حيث تخزن الوحدات الصوتية في معجم خاص، ويكون هذا المعجم مفتوحا لاستقبال أي إستعمال للكلمات الأخرى، فتحصل حينئذ المعالجة الحاسوبية، بأن تعالج كل واحدة منها بوضع علامات على الجزء الثابت في كل من الصوتين اللذين يكونان الوحدة وعلى فترات التذبذب للأصوات المجهورة[1].

(1) ينظر: سالم الغزالي، المعالجة الآلية للكلام المنطوق: التعرف والآلية، استخدام اللغة العربية في المعلوماتية، المنظمة العربية للتربية والثقافة والعلوم، تونس، [د.ط]، 1996م، ص 72.

وبعد الانتهاء من عملية تخزين الوحدات الصوتية ومعالجة الكلام المنطوق، فإن معالجة الكلام المكتوب أيسر، وتفيد حوسبتها في البعد الاتصالي من جهة، وفي خدمة اللغة العربية والحفاظ على هويتها الثقافية من جهة أخرى[1].

وفي سبيل معجم آلي عربي؛ تحتاج النصوص العربية إلى الحوسبة/ المعالجة في سبيل معجم تاريخي، وفي سبيل معجم حضاري، يمثل الحاسوب دورا رئيسا في المعجمية الحديثة، وخاصة فيما يتعلق بحفظ مراحل التطور التاريخي لمعاني المفردات ... كونه الملاذ الأخير، لمعالجة اللغة العربية تحليلا وإنشاءً، معجما جامعا يتجاوز المعجم التاريخي إلى أن يكون معجما حضاريا يرصد حياة اللغة وحياة ألفاظها عبر دورات الزمان والمكان...، وفي عصر المعلوماتية لا بد من تزويد المعجم العربي بالوسائط المتعددة، فيزدان بالصور والرسوم التي تساعد على الشرح وتنقل المعاني بدقة وأمانة[2].

وفي هذا السياق يمكن لنا أن نمثل بنموذج لمعجم آلي عربي وهو معجم "الذخيرة اللغوية"؛ الذي يعدُّ بنكا آليا ومحوسبا من النصوص القديمة والحديثة، تشتمل على التراث العربي والعالمي، وعلى الإنتاج العلمي العربي والعالمي، وتهدف إلى[3]:

- أن تكون قاعدة معطيات نصية (مدونة لغوية)؛ يحرر على أساسه عدد من المعاجم وخاصة المعجم الجامع للغة العربية.
- أن تكون مصدرا موثوقا من المعلومات العلمية والتكنولوجية والأدبية واللغوية والاجتماعية والثقافية عامة، موجهة للباحثين والدارسين والمهتمين بهذا المجال على حدّ سواء.
- أن تستخدم كأداة تعليمية وتربوية في تعليم اللغة العربية خاصة، واللغة الأجنبية عامة.

---

(1) ينظر: عبد الله أبو هيف، مستقبل اللغة العربية، مرجع سابق.

(2) بتصرف من كتاب: جوزف طانيوس لبس، المعلوماتية واللغة والأدب والحضارة، المؤسسة الحديثة للكتاب، لبنان، ط1، 2011م، ص 37. عن: نهاد الموسى، العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، ص: 285-286.

(3) ينظر: عمر مهديوي، دور قاعدة المعطيات في بناء المعجم الإلكتروني العربي، الصوتيات، حولية أكاديمية محكمة تصدر عن مخبر الصوتيات العربية الحديثة، جامعة سعد دحلب، البليدة-الجزائر، ع: 12، أفريل 2012م، ص 24 و25.

خاتمة:

وفي ختام هذا البحث، نؤكد أن عملنا هذا جاء مساءلة لمجموعة من الأبحاث والدراسات، وما جاء فيه أفكار تظل محاولة قاصرة تتلمّس الطريق إلى تدعيم نتائجه التي نحن بصدد استكمالها في ضوء عدد من الجهود التي تصب في مجال الحوسبة الآلية للمعجم العربي، وأهم ما توصل إليه البحث من نتائج:

- يعد المعجم بؤرة الأنظمة اللغوية المختلفة، ونقطة انطلاقة للدراسات اللغوية في مختلف أصعدتها، بما في ذلك مجال الحوسبة.
- إنَّ الحركة المعجمية العربية التي شهدها التاريخ منذ قرون مضت لها الأفضلية في التفوق كمًّا وكيْفًا، على غرار ما أحرزته الأمم الأخرى في مجال المعجمية، والتراث المعجمي العربي خير شاهد على أسبقيته لما يحتويه من مؤلّفات معجمية ضخمة باختلاف مناهجها وأنماطها.
- يعدّ المعجم العربي القديم اللبنة الأساسية لبناء المعاجم العربية الحديثة، والمدونة الأساسية لها.
- يعدّ الترابط العضوي بين الحاسوبية واللغة العربية في تقديم توصيف شامل ودقيق للنظام اللغوي للحاسوب تمكّنه من مضاهاة الإنسان في كفايته وأدائه اللغويين.
- يلعب الحاسوب دورا هامّا في صناعة المعاجم الآلية؛ وذلك في احتوائه على ذاكرة هائلة تمكّنه من تخزين عدد كبير من المعلومات، وترتيبها، وتسهيل عملية البحث عنها واسترجاعها، وتحليل العلاقات بين مفردات المعجم.
- من أجل بناء معجم حضاري يواكب ما نحن فيه من عصر المعلوماتية والرقمنة الآلية، لابدّ من جمع المادة وترتيبها وضبط مداخلها ومعالجة معانيها، وتنيظم كل هذه في قاعدة معطيات وأيقونات محوسبة خاصة بالمعجم، ليكون مصدرا موثوقا يكتسي حلة المعلوماتية والرقمنة الآلية في شتى مناحي الحياة، موجها للباحثين والدارسين على حدٍّ سواء.

- للمعجم العربي بنوعيه: (الـورقي والالكترونـي) دور مهـم في تحـصيل الثـروة اللغويـة وإنمائها وإثرائها لدى الباحثين والدارسين على اختلاف مستوياتهم.

ورغم التحديات التي تواجه سبيل حوسبة المعجم العربـي؛ فإننـا نوصـي بجملـة مـن المقترحات التي نراها كفيلة بتطوير وإثراء هذا الحقل العلمي:

أولها: أن تتضافر الجهود (الفردية والجماعية والمؤسسات والهيئات) في مجال الـصناعة المعجمية الحاسوبية العربية خاصة بين اللسانيين والحاسوبيين والمهندسين على السّواء.

وثانيها: إنشاء مراكز ومعاهد وأقسام خاصة بمجال اللسانيات الحاسوبية عامة ومجـال المعجمية الحاسوبية خاصة.

وثالثها: وضع آليـات كفيلـة للنفـاذ إلى معجـم آلـي تفـاعلي في شـتى التخصـصات للباحثين والدارسين.

قائمة المصادر والمراجع:

المعاجم العربية:

1- أنطوان نعمة وآخرون، المنجد في اللغة العربية المعاصرة، دار المشرق، بيروت-لبنان، ط2، 2001م.
2- تأليف وإعداد جماعة من كبار اللغويين العرب، المعجم العربي الأساسي، المنظمة العربية للتربية والثقافة والعلوم، لاروس-باريس، 1989م.
3- اللجنة الدائمة للرصيد اللغوي، الرصيد اللغوي الوظيفي للمرحلة الأولى من التعليم الابتدائي، تونس، [د.ط]، 1976م.
4- مجمع اللغة العربية، المعجم الوسيط، القاهرة-مصر، ط3، 1985م.
5- محمود إسماعيل صيني، وحيمور حسن يوسف، معجم الطلاب - معجم سياقي للكلمات الشائعة، مكتبة لبنان، بيروت، [د.ط]، 1991م.

الكتب العربية:

6- أحمد الشرقاوي إقبال، معجم المعاجم تعريف بنحو ألف ونصف الألف من المعاجم العربية التراثية، دار الغرب الإيلامي، بيروت-لبنان، ط1، 1987م.
7- جوزف طانيوس لبس، المعلوماتية واللغة والأدب والحضارة، المؤسسة الحديثة للكتاب، لبنان، ط1، 2011م.
8- حسين نصار، المعجم العربي: نشأته وتطوره، دار مصر للطباعة، مصر، ط2، 1956م، ج1.
9- حكمت عشلي، المعجم العربي في لبنان، دار ابن خلدون، لبنان، [د.ط]، 1982م.
10- سالم الغزالي، المعالجة الآلية للكلام المنطوق: التعرف والآلية، استخدام اللغة العربية في المعلوماتية، المنظمة العربية للتربية والثقافة والعلوم، تونس، [د.ط]، 1996م.
11- صلاح راوي، المدارس المعجمية العربية: نشأتها - تطورها - مناهجها، دار الثقافة العربية، القاهرة-مصر، [د.ط]، 1990م.

12- عدنان الخطيب، المعجم العربي بين الماضي والحاضر، مكتبة لبنان ناشرون، بيروت- لبنان، ط1، 1994م.
13- نبيل علي، اللغة العربية والحاسوب، مؤسسة التعريب، الكويت، [د.ط]، 1988م.
14- نهاد الموسى، العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، المؤسسة العربية للدراسات والنشر، لبنان، ط1، 2000م.
15- وليد إبراهيم الحاج، اللغة العربية ووسائل الاتصال الحديثة، دار البداية ناشرون وموزعون، الأردن، ط1، 2007م.

المجلات والدوريات:

16- أحمد محمد المعتوق، المعجم الآلي: أهميته وسبل مقاربته بالمعجم الورقي، مجلة الفيصل، ع: 290، 2000م.
17- أحمد هاشم أحمد السامرائي، حوسبة المعجم العربي والقضايا المعاصرة، مجلة سر من رأى، ع: 34، م: 9، السنة التاسعة، تموز 2013م.
18- دائرة الحوار، اللغة العربية والحاسوب، مجلة عالم الفكر، ع: 2، يوليو- أغسطس - سبتمبر 1988م.
19- راضية بن عريبة، حوسبة النظام اللغوي العربي، مجلة اللغة العربية، الجزائر، ع:37.
20- عبد الله أبو هيف، مستقبل اللغة العربية: حوسبة المعجم العربي ومشكلاته التقنية أنموذجا، مجلة التراث العربي، دمشق، ع: 93 و94، السنة الرابعة والعشرون، آذار وحزيران 2004م/ المحرم وربيع الثاني 1424هـ.
21- عمر مهديوي، دور قاعدة المعطيات في بناء المعجم الإلكتروني العربي، الصوتيات، حولية أكاديمية محكمة تصدر عن مخبر الصوتيات العربية الحديثة، جامعة سعد دحلب، البليدة-الجزائر، ع: 12، أفريل 2012م.
22- محمد الحناش، مشروع نظرية حاسوب –لسانية في بناء معاجم آلية للغة العربية، مجلة التواصل اللساني، ع:2، م: 2، 1990م.

23- مروش ذياب زغدودة، علاقة اللغة العربية بالحاسوب، مجلة الآداب والعلوم الإنسانية، جامعة الحاج لخضر باتنة -1-، الجزائر، ع: 3، ديسمبر 2009م.

المؤتمرات والندوات:

24- عمر مدكور، المعجم بين الورقية والحاسوبية، مقال منشور في كتاب أعمال المؤتمر الدولي الأول حول: (اللغة العربية وآدابها: نظرة معاصرة)، قسم اللغة العربية، جامعة كيرالا، الهند، 2015م، ص 179.

25- محمود فهمي حجازي، المعجمات العربية وموقعها بين المعجمات العالمية، بحوث ندوة خاصة بمناسبة الانتهاء من تحقيق وطباعة معجم تاج العروس، المجلس الوطني للثقافة والفنون والآداب، الكويت، يومي: 09 و10 فبراير 2002م.

26- مروان البواب، منهج إعداد معجم العربي الحاسوبي للغة العربية، الاجتماع الثاني لخبراء المعجم الحاسوبي التفاعلي للغة العربية، 1429هـ/ 2008م.

المواقع الإلكترونية:

27- باسم أغبر، علم اللسانيات الحاسوبية: تعريفه ومجالاته، مقال منشور في موقع: مدونات الجزيرة، على الرابط التالي: blogs.aljazeera.net، يوم: 3/ 7/ 2017م.

28- مروان البواب، أثر التقانات الحديثة في تجديد المعجم العربي، مقال منشور في موقع: www.alukah.net، يوم: 18/ 03/ 2009م.

Authorship Verification in Arabic Using Function Words: A Controversial Case Study of Imam Ali's Book Peak of Eloquence

Prof. Khalid Shakir Hussein (PhD)
Thi-Qar University, College of Arts, English Department

إلتحَقق من التأليف في العربية بأستخدام الكلمات الوظيفية

(دراسة حالة مثيرة للجدل حول عائدية تأليف كتاب نهج البلاغة للإمام علي)

أ.د. خالد شاكر حسين
جامعة ذي قار, كلية الآداب, قسم اللغة الإنكليزية

Abstract: This paper addresses the viability of two multivariate methods (Principal Components Analysis and Cluster Analysis) in verifying the disputed authorship of a famous Arabic religious book called (Nahjul-Balagha/ Peak of Eloquence). This book occupies an exceptional position in the history of the huge debates held between the two basic Islamic sectors: Sunni'e and Shia. Therefore, it represents a serious challenge to the viability of the multivariate techniques in resolving certain types of historical and sectarian conflicts and controversies. Furthermore, verifying the authorship of this book could be a good opportunity to find out whether there are certain quantitative techniques of attribution that hold for different languages such as English and Arabic. Function words have been targeted in this paper as possible indicators of the author's identity. Accordingly, a set of Arabic function words would be tested using WordSmith Tools (version 5). It turned out that the multivariate techniques are most likely robust for addressing the type of issues raised about Nahjul-Balagha. Besides, it appeared

that the statistical patterns of function word usages are quite sensitive to genre in Arabic.
Keywords: Authorship Verification, Arabic Language Processing, Stylometrics,

المستخلص:

يتناول هذا البحث القابلية التطبيقية لمقاربتين ذات متغيرات متعددة هما (تحليل المركبات الأساسية والتحليل العنقودي) ودورهما في التحقق من قضية التأليف المتنازع لكتاب ديني ذائع الصيت في العربية يدعى (نهج البلاغة). يحتل هذا الكتاب مكانةً استثنائية في تأريخ المناظرات الهائلة بين الطائفتين الأسلاميتين: السنة والشيعة. لذا يمثل هذا الكتاب تحدياً جاداً للقابلية التطبيقية الخاصة بالمقاربة التي تعتمد متغيرات متعددة ودورها في حل النزاعات والخلافات التأريخية والطائفية. فضلاً عن ذلك, قد يوفر التحقق من تأليف (نهج البلاغة) فرصةً سانحةً للكشف عن وجود بعض المقاربات الكمية لأسناد التأليف والتي تكون متوفرة لمختلف اللغات مثل العربية والإنكليزية. تناول هذا البحث الكلمات الوظيفية بصفتها مؤشرات مُرَجَحة تُشير الى هوية المؤلف. ولذا تفحص الباحث مجموعة من الكلمات الوظيفية مستخدماً الأدوات المتوفرة في برنامج (أدوات ووردسميث / النسخة الخامسة). تبين أن المقاربات التي تعتمد متغيرات متعددة ذات جدوى عالية الترجيح في التعاطي مع نوع القضايا التي تُثار حول (نهج البلاغة). وفضلاً عن ذلك, تبين أن الصيغ الإحصائية لأستعمالات الكلمات الوظيفية تتأثر كثيراً بنوع الجنس الأدبي في العربية.

Introduction

This paper is an attempt to test the capability and efficiency of the multivariate methods in settling down a real case study of verifying the authorship of a heatedly debated Arabic text. The book under investigation stands as a hallmark of the recently increased hassle between the two basic Islamic sectors: Sunni'e and Shia. Its authenticity is the real moot point that stimulates all kinds

of sectarian dispute across the Islamic world. Here comes the role of the quantitative attributional verification to conduct a rather objective investigation of the problem.

Authorship verification is a particular case of authorship attribution. Generally speaking, the questions usually posed in authorship analysis might vary according to the circumstances of the cases under investigation. However, two questions are basically addressed in authorship analysis: "*which author, among a clear-cut set of candidate authors, has written the questionable document?*" " . . . *Did a particular author write the document?*" (Luyckx, and Daelemans 2008). The first question is approximated by attributing the disputed texts to one of the *n* candidate authors. The second is a consistency question approximated in cases where the researcher is faced with a set of textual samples attributed to a single author and he has to figure out whether a given disputed text sample belongs to this set or not.

Nahjul-Balagha Authorship Controversy

Nahjul-Balagha (Peak of Eloquence) is a religious text well-known for its highly distinctive *genre* and *style*. The distinctive figurative style of the book and the various genres it involves crucially contribute to its uniqueness within the rhetorical map of Arabic. It comprises a collection of *sermons*, *letters*, and *sayings* attributed to Imam *Ali Ibn Abi Talib*, cousin and son-in-law of Prophet *Muhammed*. This collection was compiled in a book in the fourth century (A.H) by a well-respected Shi'i scholar and poet Shareef Razi over 300 years after Imam Ali (Nahjul-Balagha 1998). The text he had selected fall into three different genres: 241 *sermons*, 79 *letters* and 489 *sayings*. These numbers may vary in different editions of Nahjul-Balagha.

Known as a literary masterpiece, this book is deemed exceptionally distinctive in its eloquent standards and style in Shia Islam. It is seen by Shia scholars as being above the words of human beings and below the words of Allah and third only to the

Qur'an and Prophetic narrations (see http://www.nahjulbalagha.org/sermons.php).

Since the time it has been collected, Nahjul-Balagha was and still a subject of analysis and interpretation. The genuineness and authenticity of this book were not questioned by Shia scholars. They think that the references Al-Razi had made to the sources from which he collected the sermons, letters and sayings of Imam Ali were quite enough to guarantee the authenticity of the book (ibid.). However, Al-Razi did not dedicate an independent bibliography for the sources, he instead referred occasionally to certain books and their compilers as he incorporated some explanatory notes at the margins of the expounded sermons (see shttp://www.al-islam.org/nahjul/sources.htm) .

The first person that suspected the attribution of Nahjul-Balagha to Imam Ali was *Ibn Khalikan*, a Sunni scholar (see http://www.islamology.com/Resources/Nahj-Imam/main/main1.htm). Then a series of Sunni scholars followed Ibn Khalikan's doubts: Ibn Al-Athir Al-Jazari, Salah Al-Din Al-Safadi, and Al-Dhahabi. Nevertheless, all the reasons such scholars used to back up their attributional doubts are more or less ideological and biased rather than professional and technical.

The most controversial sermon that received a huge amount of denial from the Sunni scholars is called *Al-Shaqshaqiyyah* (the roar of a camel) due to its sensitive subject matter. Imam Ali in this sermon unveiled quite clear and direct indications of the *Caliphate* being snatched from him by the two caliphs *Abu-Baker* and *Umar*. Most Sunni scholars abhor the downgrading of the two caliphs explicitly expressed throughout this sermon. Therefore, this sermon, in particular, was and still deeply doubted and repeatedly attacked by them and seen as being forged and unauthentic. Some other doubts are raised nowadays on Websites by some modern *Wahhabi* scholars claiming that the whole book has been authored collaboratively by a number of Shia orators and outspoken experts of literature (see www.sunniforum.com/what do Sunnis say about

Nahjul-Balagha).

The researcher will apply two *multivariate techniques* (Principal Component Analysis and Cluster Analysis, henceforth PCA and CA) to address only two types of skeptical questions commonly raised about the authorship attribution of Nahjul-Balagha. First, *does this collection of sermons, letters, and sayings share a single author?* Or *is it the product of a multiple-author conspiracy?* Second, *does the sermon "Al-Shaqshaqiyyah" belong to this collection?*

Methodology

Six methodological procedures will be followed in analyzing Nahjul-Balagha corpus:

1- Since the samples selected for this study are machine-readable, the scanning or retyping processes could be a very threatening source of all types of errors. The researcher tried his best to check the authenticity of each sample making sure that each one is highly representative of the hard copy. What is more, all the non-authorial materials have been removed from the main body of the texts such as, *titles of sermons*, *author names*, *dates*, *Quranic verses*, *Prophetic narrations*, poetic lines, etc

2- transcribing every individual sample into *plain text format*

3- incorporating all the samples into one master corpus

4- analyzing samples with their master corpus via *WordSmith Tools (5.0)* for frequency and word count, besides producing some sort of charts representing basic statistical descriptions

5- importing WordSmith Tools (5.0) outputs into an excel spreadsheet in a form of matrix

6- conducting a thorough statistical analysis to the matrix using SPSS (14.0) (Principal Components Analysis/PCA and Cluster Analysis/CA).

Function Words in Arabic

Function words might be the commonest features that have been counted in the computational studies of authorship attribution. Ever since Mosteller and Wallace (1964) published their influential study of the Federalist Papers, function words have been the focus of interest in numerous papers and studies dealing with these words as possible indicators of authorial styles. Burrows (1987) has conducted another pioneeric study of function words demonstrating the efficacy of such words for attributing different texts and samples to different authors. Then, Grieve (2007) has produced an extensive quantitative evaluation of attributional techniques that underscored the powerful performance of function words in discriminating various authorial styles.

The appeal of function words in attributional studies lies in their being important markers of authorial individuality. Much has been written about the rationale behind the assumption that people tend to express themselves in stable and unique patterns of function words usage. This rationale almost always instigates three salient characteristics about function words: *their high frequency, low semantic load*, and the very fact that *their usage lies beyond our conscious control* (Zhao and Zobel 2005: 174-189).

Determining *function words* hierarchical lists in Nahjul-Balagha, however, is not an easy process. Arabic morphology is quite complex and demanding when it comes to the morphology of function words. There is a diversity of suffixes and prefixes that should be detached from the basic function words before measuring their frequencies and distributions. The inflectional morphology of Arabic created a dilemma for the researcher as to whether to target only the function words occurred distinctly stripped off any further affixational elements, for example (Min/from; Fi/in; Ina/is; . . . etc.) or to target the distinct ones along with their affixational occurrences, for example (Min/from- Min*aa*/from us- Min*hu*/from him- Min*ha*/from her- Min*huma*/from them (dual)- Min*huna*/ form them (femi.)- Min*ki*/from you (femi.), Min*ka*/from you (masc.))

It was a crucial choice to be made, though the first choice sounds too easy to be worked out. On the one hand, selecting the non-affixational function words would save much time in figuring out the wordlist and even in conducting the statistical analyses. On the other hand, if we skip the affixed function words there might be a high risk of disturbing the actual statistics of these linguistic items leading to probably serious problems in attribution.

It would be more reasonable if we try our hands at both choices to see how far the effects would be on the attributional process. First, the researcher will consider Arabic function words that occur distinctly with no affixes whatsoever. Then the function words will be accounted for, in the second choice, regardless of how many affixes are attached to them. Nevertheless, the traditional basic statistics of the book will be explored first to find out how far the traditional statistical characteristics might be of use in resolving the questions raised in this paper.

Results (Basic statistics)

The collected texts in Nahjul-Balagha have been segmented into eleven (5,000) token blocks or samples. The authentication of these samples was painstaking due to the many and various citations incorporated into the textual body of the samples in question. The citations removed by the researcher were a diversity of Quranic verses, poetic lines, major titles, Prophetic narrations and alike. Then the eleven samples were grouped into one master corpus. The latter was subjected to hierarchical processing for frequency and the table (1) below explores the basic statistical descriptions of the master corpus.

The mean of *word-length* is in particular appealing in the table below. The extremely narrow range of the word-length means (from 7.48 to 8.13) indicates a possibly significant characteristic of a single authorship. As for the mean of *sentence-length* in words, it is really statistically turbulent with a wide numerical band (from 22.80 to 34.81). There are (twelve) figures fluctuating between the

statistical profiles of the samples. Therefore, there is a rather wide range of statistical transition. Hence, the mean of sentence-length does not help much in drawing any significant conclusion regarding the authorship of Nahjul-Balagha.

Table 1. Basic Statistics of Nahajul-Balagha Corpus

	1	2	3	4	5	6	7	8	9	10	11
File Size	45,145	46,039	46,862	45,882	45,930	45,398	47,406	45,681	44,223	44,640	33,615
tokens	5,016	5,015	5,023	5,017	5,016	5,001	5,021	5,025	5,049	5,025	3,857
types	3,341	3,425	3,410	3,190	3,346	3,112	3,367	3,248	3,038	2,994	2,288
type/token ratio (TTR)	66.61	68.30	67.89	63.58	66.71	62.25	67.06	64.65	60.18	59.58	59.32
standardised TTR	76.46	77.78	76.56	74.26	76.04	72.66	76.22	73.40	71.42	70.24	69.60
standardised TTR std.dev.	18.56	17.83	18.76	21.01	20.06	23.08	19.28	24.13	22.84	23.60	23.40
standardised TTR basis	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
mean word length (in characters)	7.70	7.89	8.05	7.89	7.86	7.79	8.13	7.80	7.51	7.66	7.48
word length std.dev.	2.85	2.88	2.98	2.92	2.91	2.98	3.04	2.96	2.81	2.88	2.80
sentences	217	178	156	220	197	187	176	163	145	173	121
mean (in words)	23.12	28.17	32.20	22.80	25.46	26.73	28.53	30.82	34.81	29.05	31.88
std.dev.	19.11	21.21	25.26	16.18	18.68	20.74	22.00	25.37	26.20	20.88	20.33

The lexical diversity is indicated by a comparatively narrow range (from 59.32 to 68.30). The nine figures continuum of this range suggests an exceptionally rich vocabulary attributed to the claimed author of the samples. One particular finding, however, might be reached if one gives a quizzical look at the values of the Type/Token Ratio (henceforth TTR). There is an observable shift in the TTR values scored in samples (9), (10), (11) showing (60.18; 59.58; 59.32) respectively (see **Figure 1** below). This range of values sounds quite stable and distinct in comparison to that of the rest of samples which ranges from (62.25) to (68.30).

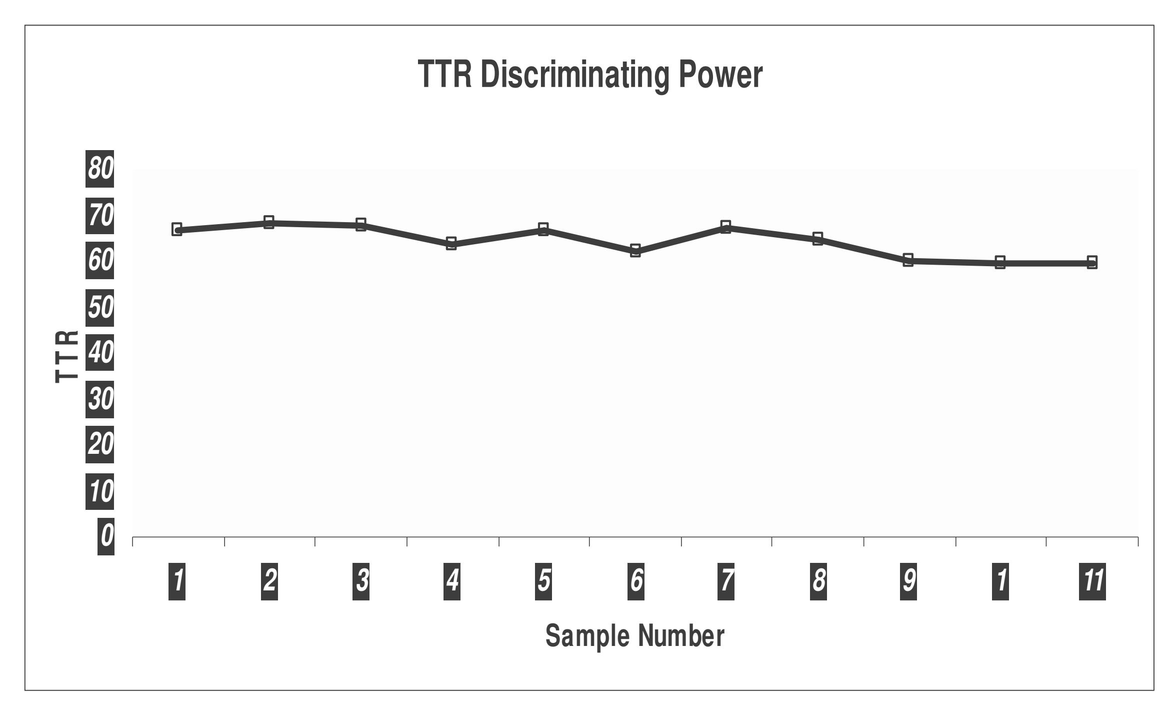


Figure 1. TTR Discriminating Power in Nahajul-Balagha Corpus

It seems that TTR values are perhaps highly sensitive to genre: the samples (9, 10) constitute the *letter-blocks*, and (11) the *saying-block* in Nahjul-Balagha, whereas the remaining ones represent the *sermon-blocks*. Nevertheless, TTR does not provide us with any decisive and finite conclusions about the two questions raised about the authorship of Nahjul-Balagha. How about the Hapax Legomena/Type Ratio (henceforth HTR)?

The contribution of the notion of *hapax-legomena* to discriminate authorial identities is deeply entrenched in their low distribution or infrequent occurrence. A hapax is commonly defined as a word that occurs only *once* in a text or a corpus (Lardilleux & Lepage, 2007: 458). The distinctive power of hapaxes comes about from the parallelism between their unique occurrence, on the one hand, and the author's linguistic uniqueness or *idiolect*, on the other.

If two texts are available for comparison, the vocabulary used in both texts can be compared *quantitatively* so that a particular "qualitative textual analysis" can be backed up (Johnson ,1997), and the amount of the shared hapaxes is supposed to be quite powerful in establishing authorial distinctiveness: *the more there is, the less the authorial differentiation.*

A high number of *shared* hapax legomena and a low number of *unique* hapax legomena are both quite useful in measuring up how likely the authorial differentiation is (cited in Coulthard et al., 2010: 526). So, what is looked for here is the particular type of the items shared. Those items shared *once only* (hapaxes) are supposed to be *insignificant* in terms of the main concern of the text. This would explain why they have been used less frequently, actually just once. Otherwise, they would occur more frequently rather than showing up for once (Coulthard, 2004: 5). Thus the hypothesis is made so clear by Coulthard (ibid.) in this concern: "the chances of two writers independently choosing several of the same words for single use are so remote as to be discountable".

What is more, the researcher thinks that there is a relatively stable ratio of hapax that can be safely and distinctively attributed to authors throughout their writings. This ratio compares the number of hapaxes to the number of *types* (not *tokens*) along the text itself. It might be a quite useful marker in raising doubts about the occurrence of plagiarism as a case of authorial indifferentiation. Statistically speaking, any extreme variation in the figures alluding to the ratio of hapaxes to types is most likely to alert suspicion of plagiarism or doubtful authorial identity.

Despite the negative attitude most statisticians have to hapaxes since they used to neglect those tokens and types attaining low frequency, hapaxes in particular constitute something like (40%) of words used in a corpus (Lardilleux & Lepage, 2007: 458). Beside the type/token ratio, the (HTR) enhances the indication of vocabulary richness. The more hapaxes are there, the richer the vocabulary. This can give a measure of the author's distinctive originality that can bring him into a sort of statistical discrimination.

The eleven samples of Nahajul-Balagha reveal insightful and noticeable harmony in the HTR values (see Table 2 below). The statistical consistency of the values provide us with an unavoidable avenue to draw rather concrete evidence of *unitary authorship* of Nahjul-Balagha. All the samples share a great deal of similar HTR values ranging from (0.85) to (0.88). **Figure 2** shows some sort of a straight line extending throughout the stable node-samples with roughly indiscriminate plottings.

Table 2. HTR Calculated for the Master Corpus of Nahajul-Balagha

Author	Segment	Types	Hapax Legomena	HTR
Imam Ali	1	3341	2926	**0.87**
	2	3425	3042	**0.88**
	3	3410	3033	**0.88**
	4	3190	2774	**0.86**
	5	3346	2958	**0.88**
	6	3112	2718	**0.87**
	7	3367	2961	**0.87**
	8	3248	2866	**0.88**
	9	3038	2629	**0.86**
	10	2994	2568	**0.85**
	11	2288	1978	**0.86**

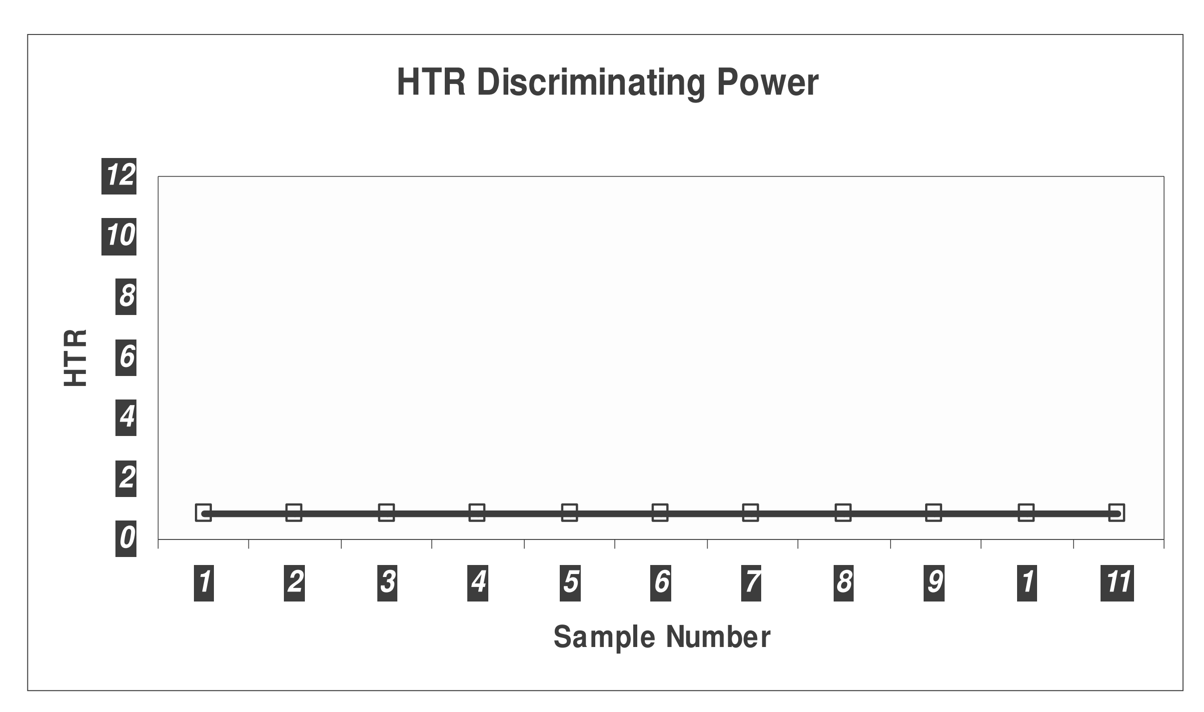


Figure 2. HTR Discriminating Power in Nahajul-Balagha Corpus

The turbulences seen in TTR values disappear in an exceptional way in the HTR plottings. This underscores the significance of HTR as a highly robust statistical feature of great usefulness in attributing samples.

Incorporating *Al-Shaqshaqiyyah* in sample (1), the HTR of this sample (0.87) stands high in favour of a strong attributional affinity that holds between this sample and the other ones. If *Al-Shaqshaqiyyah* was forged and unauthentic, it would not share the same range of unique occurrences with the other sermons. This might well be considered a plausible answer for the second question raised about Nahjul-Balagha in this study.

Distinct VS Affixed Function Words Frequency

Going through the distinct function words frequency, more than 70 function words were tabulated after processing the master corpus. However, some function words had zero frequency for at least one of the eleven samples. Thus, they were removed from the two lists. Each list was left with only 40 function words with non-zero frequencies all over the samples. Below are two tables

showing the lists of *distinct* and *affixed* function words, their distributions and percentages all through the master corpus.

Table 3. The Top-40 Distinct Function Words of Nahjul-Balagha-Corpus

WordSmith Tools -- 14/7/2013			
N	Word	Freq.	%
1	و	8509	19.0826807
2	لاَ	1505	5.657052803
3	مِنْ	1008	1.608604789
4	فِي	898	1.433062553
5	مَا	760	1.171345115
6	عَلَى	683	1.089957356
7	مَنْ	383	0.587268412
8	إِلَى	368	0.543668315
9	لَمْ	320	0.510668159
10	قَدْ	316	0.504284799
11	أَنْ	304	0.48513478
12	عَنْ	250	0.398959517
13	إِلاَّ	234	0.37342611
14	حَتَّى	190	0.303209245
15	أَوْ	178	0.284059167
16	كَانَ	170	0.279257264
17	إِنَّ	157	0.250546575
18	ذلِكَ	156	0.248950735
19	الَّذِي	136	0.217033982
20	ثُمَّ	132	0.208650623
21	بَعْدَ	117	0.200582613
22	بِمَا	106	0.181920611

23	لَوْ	96	0.169158831
24	هُوَ	94	0.156392127
25	عِنْدَ	93	0.153200448
26	بَيْنَ	93	0.153200448
27	لَيْسَ	91	0.150008783
28	هذَا	90	0.148221263
29	إذًا	84	0.143625423
30	أَمَّا	83	0.138050399
31	فِيما	79	0.128050399
32	غيرَ	72	0.122454559
33	كُلَ	71	0.121362219
34	معَ	62	0.114900343
35	أنتَ	53	0.097983576
36	قبلَ	52	0.095387743
37	إذ	51	0.092270252
38	أنا	50	0.088845017
39	التي	42	0.064258496
40	كيف	32	0.058258496

Table 4. The Top-40 Affixed Function Words of Nahjul-Balagha-Corpus

WordSmith Tools -- 14/7/2013			
N	Word	Freq.	%
1	و	8509	19.0826807
2	لاَ	1505	5.657052803
3	مِنْ	1383	5.308604789
4	فِي	1056	2.433062553
5	مَا	760	1.171345115
6	عَلَى	725	1.089957356
7	إِلَى	549	0.587268412
8	إِنَّ	392	0.548968315
9	مَنْ	383	0.543668315
10	عَنْ	346	0.510668159
11	لَمْ	320	0.504284799
12	قَدْ	316	0.48513478
13	أَنْ	304	0.398959517
14	إِلاَّ	234	0.37342611
15	كَانَ	211	0.303209245
16	حَتَّى	190	0.298059167
17	أَوْ	178	0.284059167
18	كُلِّ	168	0.248950735
19	ذلِكَ	156	0.217033982
20	الَّذِي	149	0.210650623
21	ثُمَّ	132	0.169158831
22	غَيْرِ	125	0.156392127
23	بَعْدَ	106	0.153200448
24	بَيْنَ	104	0.150008783

25	هذَا	101	0.148412943
26	هُوَ	98	0.145221263
27	أَنْتَ	97	0.143625423
28	بِمَا	96	0.141050399
29	لَوْ	94	0.132454559
30	عِنْدَ	93	0.130454559
31	لَيْسَ	90	0.128454559
32	إِذَا	83	0.114900343
33	أَمَّا	83	0.114900343
34	فِيَما	72	0.083175255
35	مَعَ	63	0.077983576
36	قَبْلَ	58	0.071387743
37	إِذْ	50	0.065270252
38	أنا	49	0.05450172
39	الَّتِي	41	0.054258496
40	كَيْفَ	32	0.051258496

The distinct function words in Table (3) above account for 37.64% of all the words in the master corpus. This percentage still falls within the limits set by Burrows' guidelines (1992), but it is quite expected that the percentage will go a little bit down after removing 30 function words with zero frequencies. Since the paper investigates the attributional potential of only function words, all content words have been removed from the lists. Therefore, the two lists ignore three content words: *Allah*- 901 tokens, 4th word on the list; *Adduniah/ the world*- 155 tokens, 27th word on the list, and *Alhaq/the truth*- 66 tokens, 70th word on the list.

However, the affixed function words in Table (4) attain comparatively a considerable percentage: they account for 42.35% of the total words in the master corpus. This percentage is soundly

higher than that of the distinct function words, 37.64%, a matter which enhances the credibility of the scored affixed words.

Analysis Matrix

The researcher designed the analysis matrix that will encompass the frequencies scored for the (40) function words. Below is a sample of the matrix, the first four words in the hierarchy as well as the segment length. The complete matrix can be found in the Appendix. The segment length is used in the matrix because the textual body can not be evenly divided into (5,000) word segments.

Sample Matrix – Nahjul-Balagha Corpus

Author	Text	Segment	Wa	La	Min	Fi	Segment Length in Tokens
A	NB	1	790	131	110	87	5,016
B	NB	2	835	111	115	96	5,015
C	NB	3	843	133	132	92	5,023
D	NB	4	762	145	118	82	5,017

Statistical Analysis through PCA

Principal components analysis was used to deal with the over-dimensionality of the 40 function words that should be reduced into a more manageable number of components (factors). It is usually the first two dimensions or components that sufficiently explain the most interesting variables and determine which variable is correlated more highly with one component rather than with another (see Craig and Burrows 2001). This is the only way by which the researcher can draw conclusions about the behavior of the distinct and affixed function words all through the eleven samples. Some function words are expected to stand out constituting salient markers with crucial role in attributing Nahjul-Balagha.

Conducting PCA on the corpus of the eleven samples indicates that the first two components could capture 42.70% of the total variance of the distinct function words within the data environment. These two components are quite enough to determine which variables are highly correlated with each one of them. The first component accounted for 23.30% of the total variance and the second factor 19.40%. Table (5) below explains the percentage of variance captured by each component.

The behavior of the affixed function words in the scattorplot below (**Figure 4**) shows slightly different statistical patterns in terms of the first two principal components accounting for 45.83% of the total variance (see Table 6 below).

Table 5. Total Variance Explained for Distinct Function Words

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.320	23.301	23.301	9.320	23.301	23.301
2	7.763	19.406	42.707	7.763	19.406	42.707

Table 6. Total Variance Explained for Affixed Function Words

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.881	23.525	23.525	9.881	23.525	23.525
2	9.373	22.317	45.843	9.373	22.317	45.843

Then, the researcher can apply PCA one step further to check the way the distinct and affixed function words behave in respect of the two principal components specified above. **Figure 3** and **4** below plot the statistical behavior of the distinct and affixed function words throughout the eleven samples of Nahjul-Balagha.

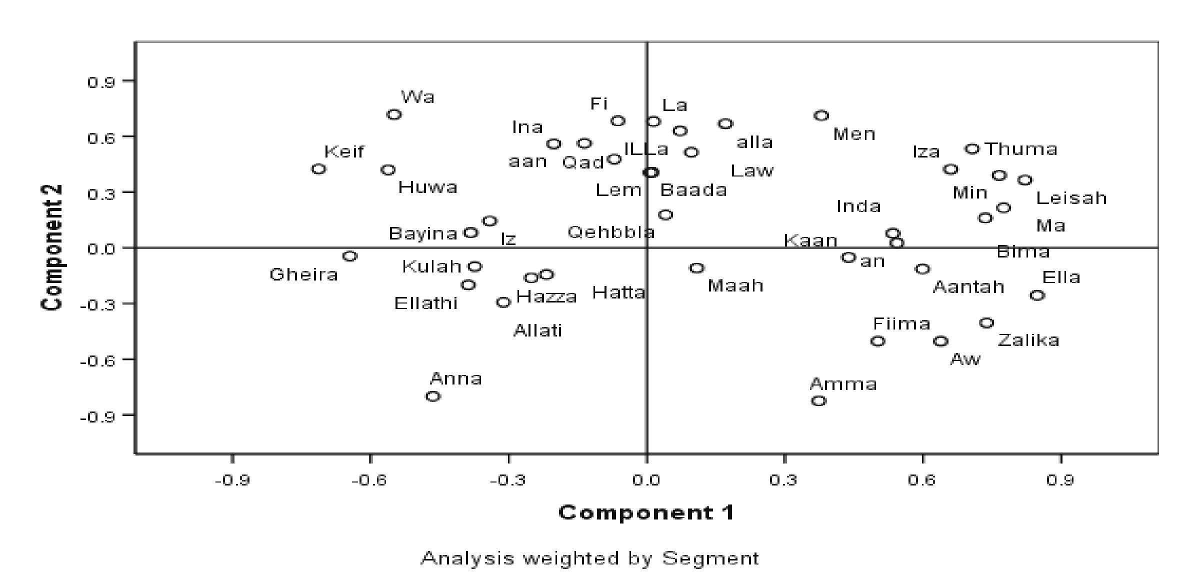


Figure 3. The 40 distinct function words behavior

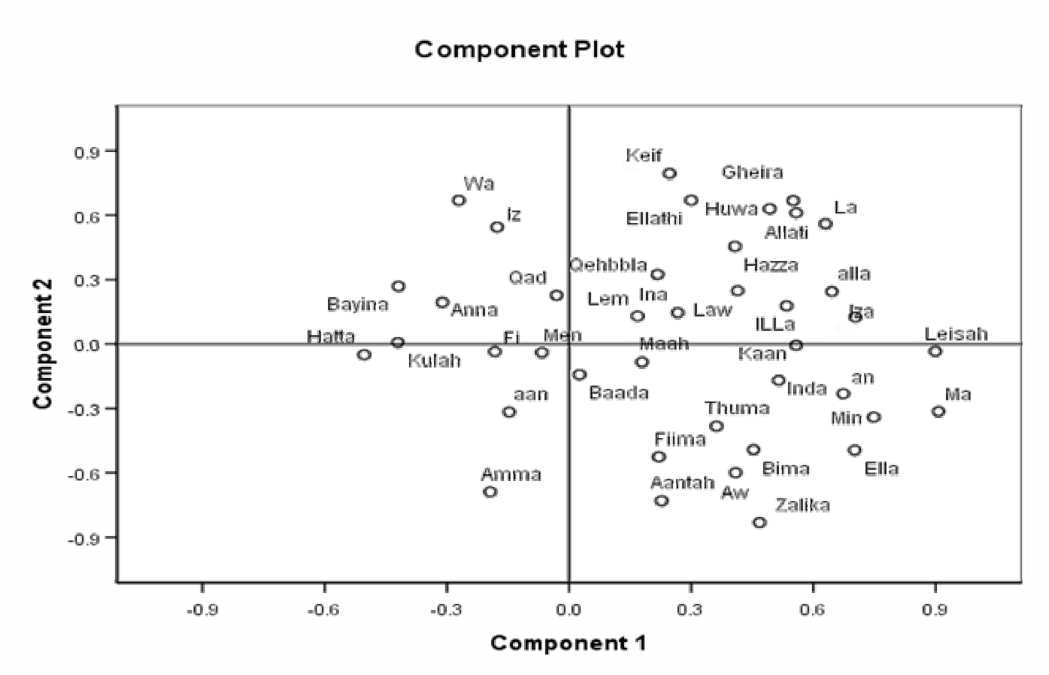


Figure 4. The 40 affixed function words behavior

The first two principal components are represented in figures **3** and **4** above by the x-axis and y-axis. The working principle is rather simple and attractive: those variables or function words that come close together tend to have a similar plotting behavior. Moreover, they tend to be found in one group of samples more

often than in another. The function words in figures **3** and **4** are not treated equally, only those variables that exist at either end of the two axes will be given an authorial weight in controlling the markers of the authorial style.

It is evident that the function words found on the far right of **Figure 3** are Ella/to, Leisah/not, Min/from, Zalika/that, Bima/as. These words together with those located on the far left of the figure (Keif/how, Gheira/but, Huwa/he, and Wa/and) are highly correlated with component (1).

As for component (2), its top is occupied by *Wa/and, Men/who, Fi/in, La/no* and *Alla/on* and the function words *Amma/either, Anna/I am, Aw/or,* and *Fiima/while* move significantly towards the bottom of it.

As for **Figure 4**, the function words *Hatta, Kulah,* and *Bayina,* on the far left of the same figure, and *Leisah, Ma,* and *Min*, on the far right, are clearly correlated with component (1). Component (2) is headed at the top by *Keif, Wa,* and *Huwa*, and at the bottom we find *Amma, Zalika,* and *Aantah*.

Tables (7) and (8) below indicate how much the variance in components (1) and (2) is conditioned by distinct and affixed function words.

Table 7. Components Matrix of PCA for Distinct Function Words

	Component	
	1	2
Wa	-.549	.718
La	.014	.681
Min	.765	.390
Fi	-.063	.684
Ma	.775	.216
Alla	.171	.668
Ella	.848	-.256
Men	.379	.712
Lem	.012	.406
Qad	-.071	.477
An	.543	.026
Aan	-.136	.563
ILLa	.072	.630
Hatta	-.218	-.144
Aw	.639	-.504
Kaan	.438	-.051
Ina	-.202	.560
Zalika	.738	-.404
Ellathi	-.388	-.200
Thuma	.707	.534
Baada	.007	.406
Huwa	-.562	.420
Bima	.735	.160
Law	.097	.514
Inda	.535	.078
Bayina	-.383	.081
Leisah	.821	.365
Hazza	-.251	-.162
Iza	.660	.423
Amma	.374	-.825
Fiima	.502	-.504
Kulah	-.374	-.101
Gheira	-.645	-.045
Maah	.108	-.109
Aantah	.599	-.114

Qehbbla	.041	.178
Anna	-.465	-.801
Iz	-.341	.144
Allati	-.312	-.293
Keif	-.713	.424

Table 8. Components Matrix of PCA for Affixed Function Words

	Component	
	1	2
Wa	-.213	.773
La	.592	.683
Min	.836	-.331
Fi	-.030	.202
Ma	.904	-.265
alla	.627	.413
Ella	.651	-.512
Men	.686	-.107
Lem	.409	.321
Qad	.254	.229
An	.633	-.151
aan	.187	-.155
ILLa	.698	.137
Hatta	-.574	-.185
Aw	.343	-.614
Kaan	.324	.163
Ina	.547	.106
Zalika	.426	-.843
Ellathi	-.087	.564
Thuma	.681	-.318
Baada	.272	.163
Huwa	.284	.756
Bima	.541	-.485
Law	.536	.428
Inda	.575	-.195
Bayina	-.479	.296
Leisah	.922	.020
Hazza	.115	.321
Iza	.819	.160

Amma	-.190	-.754
Fiima	.112	-.654
Kulah	-.487	.028
Gheira	.259	.652
Maah	.192	-.279
Aantah	.284	-.742
Qehbbla	.413	.423
Anna	-.279	-.270
Iz	.132	.613
Allati	.117	.631
Keif	-.279	.873

Plotting the statistical behavior of the 40 function words, the researcher needs to verify how much of each principal component is referenced by the individual samples. Figures **5** and **6** plot the eleven samples through a scatterplot giving us representative graphs of the way these samples behave based on 42.70% of the data variance for distinct function words and 45.84 for affixed function words.

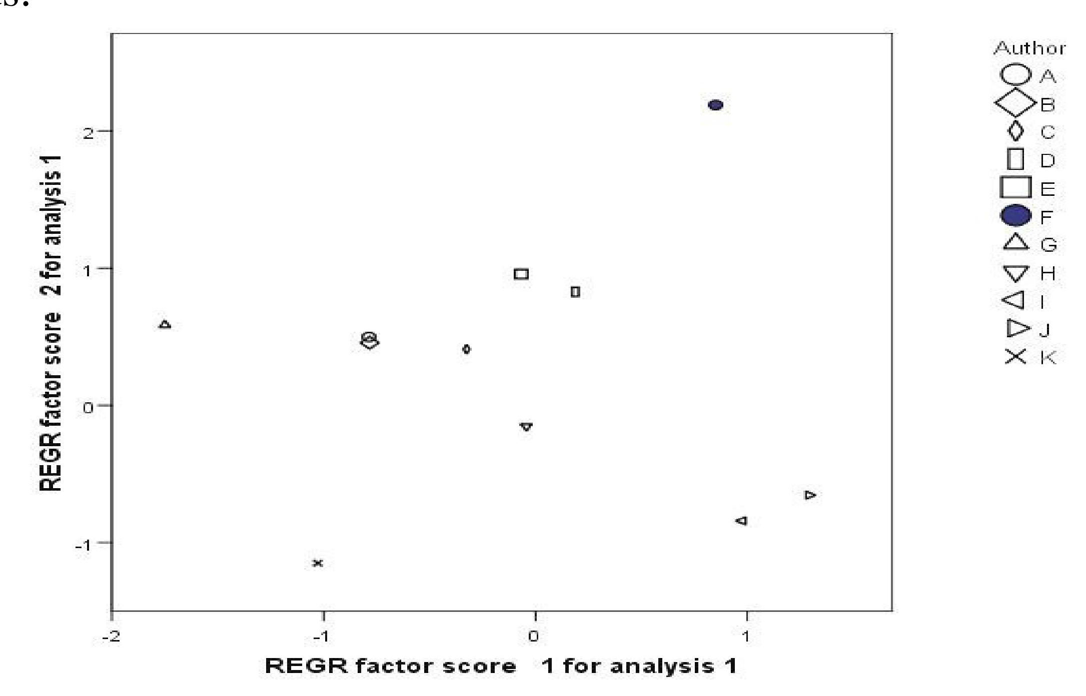


Figure 5. Scatterplot of Eleven Samples Behavior (Distinct Function Words)

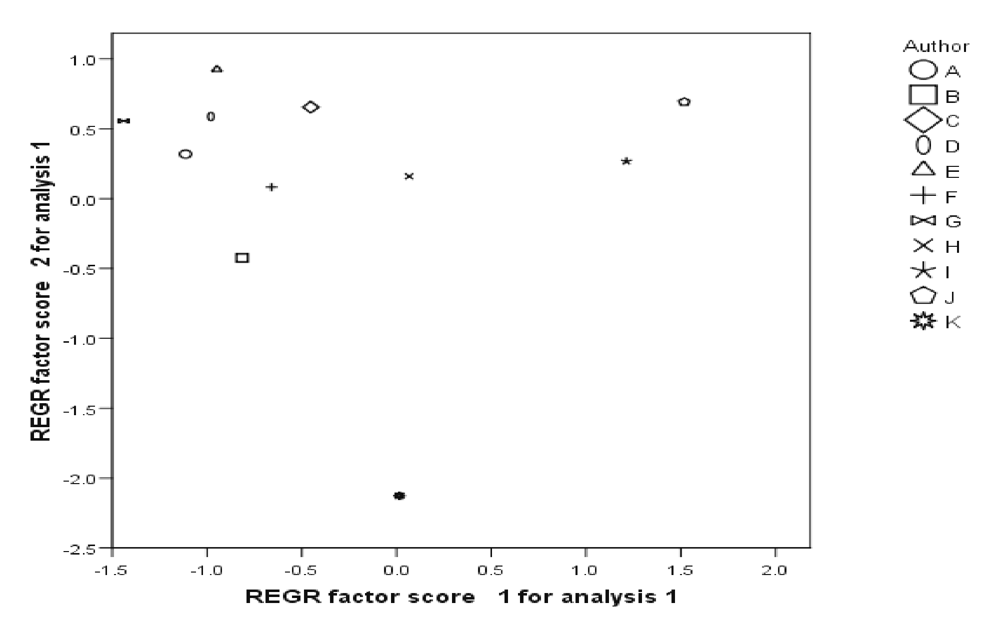


Figure 6. Scatterplot of Eleven Samples Behavior (Affixed Function Words)

The scatterplot in **Figure 5** above does not show clear cut clusters, though there is a major cluster lurking behind this chaos. It is plotted at the center of the figure with six samples orbiting around (1, 2, 3, 4, 5 and 8) [(1)] While the remaining samples (6, 7, 9, 10, 11) go in diverse and scattered directions along the two axes: (6) jumps to the top right corner, (7) diverts to the left on the further side of the horizontal axis, (9) and (10) tend to the lower right, and (11) is on the lower left.

The anomalous behavior of (6, 7, 9, 10 and 11) hints that the way distinct function words are used in these samples is inconsistent with that in (1, 2, 3, 4, 5 and 8). This could cast some serious doubts about the claimed *unitary authorship* of Nahjul-Balagha. However, the researcher would not hasten to draw any premature conclusions about the two questions raised in this case study. Further statistical analysis is required to confirm the outputs of the PCA.

Compared with the way the distinct function words behave, the vast majority of the affixed function words shown in **Figure 4**

(1) The numbers correspond to the letters A, B, C, . . .etc.

tend to display a quite notable rush to the right side of the figure. This sharp shift in behavior gives rise to dramatic repercussions over the way the individual samples behave as it could be observed in **Figure 6** above.

It is apparent by the vertical axis in **Figure 6** that samples (1, 2, 3, 4, 5, 6, 7, 8) share noticeable similarities in the patterns of using function words. This most likely consolidates a sort of stylistic harmony that made them grouped in one cluster plotted in the upper left. Also interesting to note is the way samples (9) and (10) divert from all the other samples identifying themselves in one cluster up to the right corner of the plotting. Whereas sample (11) goes down and resides as a single plot recognized on the horizontal axis.

The differences in samples behavior are dramatically associated with considerable differences in *genre*. The latter plays an undeniable role in the way the function words are used. The 241 *sermons* were segmented into eight subsequent samples represented by the first eight segments in Nahjul-Balagha corpus. Samples (9) and (10) represent the 79 *letters*. As for (11), it comprises all what is left of the 489 *sayings*. Therefore, the stylistic differences reflected by the different plottings of the eleven samples might well be attributed to the different writing styles of the three different genres: *sermons, letters,* and *sayings*. Does that mean the subconscious use of function words in Nahjul-Balagha corpus is significantly affected by the text genre?

Statistical Analysis Through CA

Hierarchical cluster analysis is used to trace possible groups that might be observed in the numerical tables that we consider for the statistical behavior of function words within the eleven samples. This analysis is always used regardless of any assumptions a researcher might make about the corpus he is interested in (Baayen 2008). Consequently, the researcher will allow this multivariate technique to trace groups or any possible clusters by calculating the

degrees of similarity or difference between all the samples under consideration.

The significance of this analysis lies in its outputs that might confirm or disconfirm the researcher's findings in PCA. The outputs are usually presented in a graphic representation that visualizes the distance at which clusters come close to each other. Distance, within cluster analysis, refers to a certain numerical taxonomy depending, in our context, on calculating the similarities and dissimilarities in the statistical and distributional behavior of function words being scored within the textual samples concerned.

The graphic representation is called *a dendrogram*. Dendrograms are usually read from left to right. The groups of membership will be allowed to form relying on the statistical characteristics of the eleven samples. No presuppositions are undertaken by the researcher, the individual samples will behave in a way that complies with their own genuine statistical features.

The groups that will be allowed to form by CA are measured in accordance with the statistical characteristics of the individual samples including the disputed sermon of (*Al-Shaqshaqiyyah*) incorporated in the body of the first sample. It is a privilege that CA does not pay attention to any assumptions might be taken about the data. It is quite independent of PCA in terms of the way the latter addresses the targeted samples. It will figure out any possible similarity calculations between the samples that PCA may have missed. The density of statistical similarity is usually indicated by *rescaled distance cluster combine* which is represented by a dotted line at the top of the dendrogram. The *rescaled distance cluster* measures the distance among samples and put it (the distance) in the form of a vertical line that groups sample-arrows depending on their numerical similarity. The nearer the vertical line the more similar the samples grouped by its arrows. Samples meeting at the first vertical line tend to be more similar than those meeting, for example, in the second or third ones. Nevertheless, CA credits or discredits the outputs produced by PCA. It should be noted that the

researcher is conducting CA assuming that he knows nothing about the clustering possibilities that the eleven samples might show.

The dendrograms reported in **Figure 7** and **8** below represent the findings of CA conducted on the eleven samples. What is immediately apparent is that there is a sort of clustering in **Figure 7** that identifies (1, 2, 3, 4, 5 and 8) rather closely with each other, though their subgroups are exhausted by some further lines at which they meet: (1) and (2) are identified most strongly on the first vertical line and then cluster with the 3rd sample. Similar clusters continue to merge indicating similarity among the individual samples: (4) and (5) cluster on the second vertical line and then align themselves with (8) on the fourth vertical line and with (7) on the fifth vertical line. The latter in turn meets with (11) on the sixth vertical line. It is notable that (9) and (10) cluster on the third vertical line, as for (6), it turns out to be an outlier clustering with little similarity to any of the samples involved.

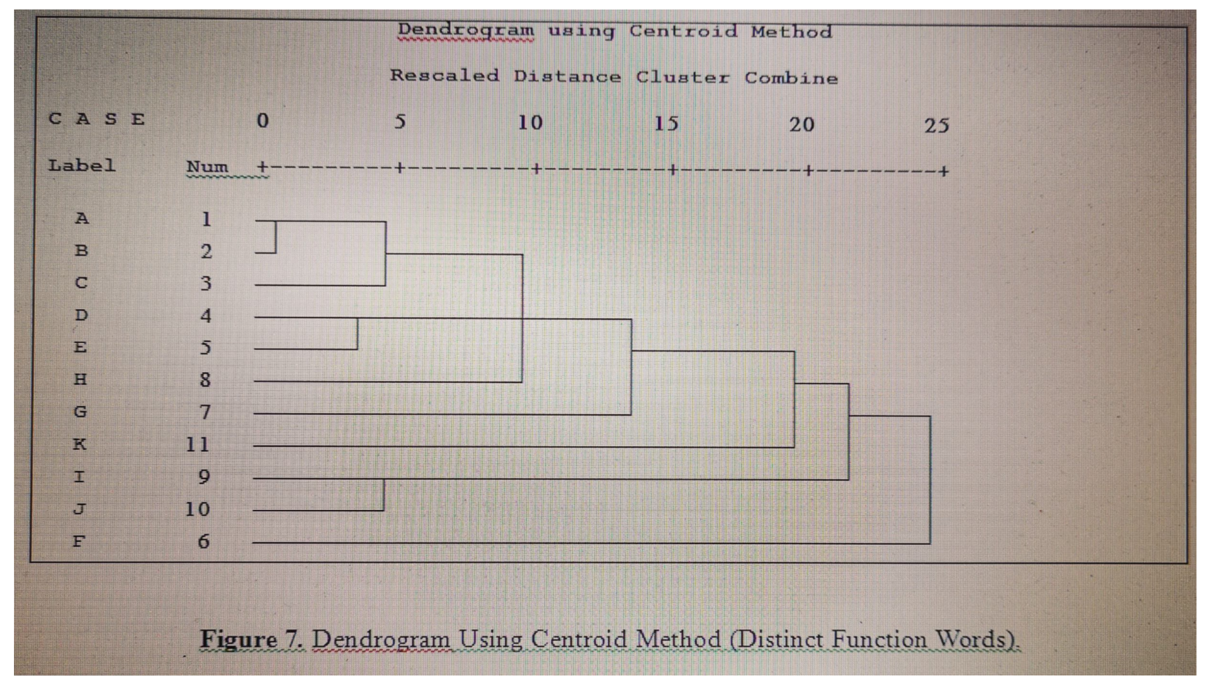


Figure 7. Dendrogram Using Centroid Method (Distinct Function Words).

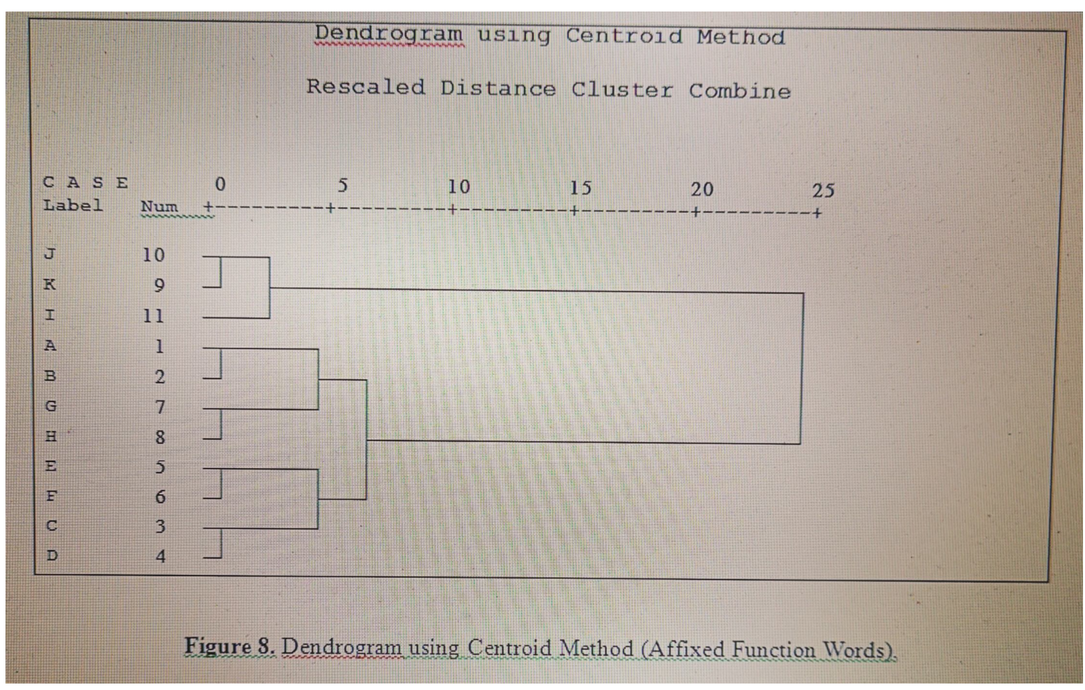


Figure 8. Dendrogram using Centroid Method (Affixed Function Words).

It looks observable that there is an *outdistancing* effect pushing the samples (6, 7, 11) in particular further into the margins of the figured analysis environment. This effect could be seen even in the ascending numbers of the vertical lines, the samples meet first at the first line and then ascend up to the sixth line. Though the overall statistical behavior of these samples does not show a quite promising uniqueness in terms of the function words usage, the researcher can build an argument based on the highly distinctive and unique cluster of the first and second samples formed on the first vertical line. If sample (2) is confidently attributed to Imam Ali, it is highly possible, then, that (1) belongs to him on the ground of the intensive degree of similarity that holds between the two samples, though this is not enough. The scattered plottings of the other samples, however, suggest an evident disapproval of attributing them to a single author.

The dendrogram in **Figure 8** represents the findings of conducting CA over the eleven samples. The first thing to be recognized in this dendrogram is the existence of two major

clusters: the first encompasses the samples (1, 2, 3, 4, 5, 6, 7, 8), and the second holds for two closely identified samples (9) and (10) with a third sample (11) laxing as an outlier. The disputed sermon included in the first sample is still strongly identified with the other samples, especially with the second one. What holds the researcher's interest is the way all the samples are identified with each other: every two samples are intensively aligned on the first vertical line. It is notable that samples (9, 10) cluster independently with a strong correlation that binds the two samples on the first vertical line.

The researcher thinks that all the statistical disharmony that hangs over the samples checked above is brought into the analysis environment by, first, ignoring those function words attached to affixes (prefixes or suffixes) and, second, by restricting frequency only to those function words stripped off any sort of affixational attachments. It is unreasonable to account for (Fi/ in), for example, only if it occurs alone as a distinct preposition. The type of information that (Fi) conveys is still stable no matter how many affixes are attached to it. Therefore, its function is lying there intact even when a series of suffixes is annexed to it: (Fi*hi*/ in it – Fi*ha*/ in her – Fi*na*/ in us – Fi*him*/ in them (masc.) – Fi*huna*/ in them (femi.) – Fi*huma*/ in them (dual)).

Even the difference in the frequencies of the 40 distinct function words listed above is too considerable to be ignored or mitigated. As one can see from the frequency list of the affixed function words below, (Fi), for example, occurs 898 times as a morphologically distinct preposition, and 1056 as an affixed preposition all through the master corpus of Nahjul-Balagha. The same can be observed for the other function words: (Min/from) rises in frequency from 1008 as a distinct to 1383 as an affixed; (Illa/to) from 368 to 549, and so on and so forth. Unless this difference in frequency is taken into consideration, the attributional statistics will not be trustworthy.

Conclusions

It turns out that the multivariate methods (PCA and CA) are perhaps robust for addressing the issues raised about Nahjul-Balagha. It is most likely that the eleven samples do share a single author. The stylistic continuities discerned throughout the stable statistical plottings and clusters of the samples would seem to suggest *unitary authorship*, even though the samples cross genre-lines.

Unlike the rest of the sermons attributed successfully only by targeting the *affixed* function words, "*Al-Shaqshaqiyyah*", interestingly, identifies itself most closely with the rest of the sermons regardless of the type of function words used whether *distinct* or *affixed*. This case study does confirm that the author of Nahjul-Balagha preserved approximately the same unique patterns of function words usage, though these patterns show a discriminatory change when crossing over genre lines (from the territory of *sermons* to that of *letters* and *sayings*). Therefore, it appears that the multivariate methods based on function words usage are highly sensitive in Arabic to genre.

Moreover, this sensitivity might surge significantly if the particular type of function words in Arabic is taken into consideration. The morphology of Arabic function words imposes unavoidable statistical repercussions that might be a source of skewed statistical data. Statistical accounts of Arabic function words could be dramatically misleading and perplexing if the particular type of function words is overlooked.

References

- Baayen, R. 2008. *Analyzing Linguistic Data*. Cambridge: Cambridge University Press.
- Burrows, J. 1987. *Computation into Criticism: A Study of Jane Austen's Novels and an experiment in Method*. Oxford: Clarendon Press.
- Burrows, J. F. 1992. "Not unless you ask nicely: the interpretive nexus between analysis and information". *Literary and Linguistic Computing*, 7(2), (91-109).
- Coulthard, M., & Johnson, A., Kredens, K., & Woolls, D. (2010). Plagiarism: four forensic linguists responses to suspected plagiarism. In Coulthard, M., and Johnson, A., *The Routledge Handbook of Forensic Linguistics*. London and New York: Routledge.
- __________ (2004). Author identification, idiolect and linguistic uniqueness. *Applied Linguistics,* 25(4), 431-47.
- raig, H.,& Burrows, J. 2001. "Lucy Hutchinson and the authorship of two seventeenth century poems: a computational approach". *The Seventeenth Century*, 16, (259-282).
- Grieve, J., 2007. "Quantitative authorship attribution: an evaluation of techniques". *Literary and Linguistic Computing*, 22(3), (251-270).
- Mosteller, F., and Wallace, D. 1984. "Inference and disputed Authorship: The Federalist". (Reading, MA: Addison-Wesley Pub. Co., 1964); 2nd edition as *Applied Bayesian and Classical Inference: The Case of the 'Federalist' Papers*. New York: Springer-Verlag.
- Ibn Abi Talib, A. 1998. *Nahjul-Balagha*, 3rd ed., Qum: Al-Thaqalein Library.
- Johnson, A., (1997). Textual kidnapping: a case of plagiarism among three student texts. *International Journal of Speech, Language and the Law,* 4(ii), 210-25.
- Lardilleux, A., & Lepage, Y., (2007). The contribution of the notion of hapax legomena to word alignment. In *Proceedings*

of the 4th Language and Technology Conference (LTC'07), 458-462.

- Luyckx, K. and Daelemans, W. 2008. "Authorship attribution and verification with many authors and limited data". *Proceedings of the 22 nd International Conference on Computational Linguistics.* Manchester: Coling, August, (513-519).
- Zhao, Y. and Zobel, J. 2005. "Effective and scalable authorship attribution using function words". In G.G. Lee et al. (Eds.): AIRS 2005, LNCS 3689 (174-189).

Appendix

Nahjul-Balagah Matrix

Author	Text	Segment	Wa	La	Min	Fi	Ma	alla	Ella	Men	Lem	Qad	an	aan	ILLa
A	NB	1	790	131	110	87	51	53	32	56	19	38	36	30	22
B	NB	2	835	111	115	96	64	61	31	35	43	26	32	28	20
C	NB	3	843	133	132	92	65	68	44	41	45	33	27	31	25
D	NB	4	762	145	118	82	68	76	38	34	24	41	41	27	41
E	NB	5	776	136	128	95	64	72	33	34	37	27	27	43	24
F	NB	6	848	199	133	90	80	75	44	25	33	28	53	18	24
G	NB	7	931	142	111	134	64	57	31	30	19	28	24	42	25
H	NB	8	750	121	136	103	67	76	41	28	29	28	49	40	18
I	NB	9	691	146	132	96	88	57	47	36	35	38	67	33	25
J	NB	10	719	151	149	110	82	80	48	43	23	23	38	31	30
K	NB	11	564	90	119	71	67	50	45	21	19	21	36	23	17

Hatta	Aw	Kaan	Ina	Zalika	Ellathi	Thuma	Baada	Huwa	Bima	Law	Inda	Bayina	Leisah	Hazza	Iza	Amma	Fiima
34	18	11	48	4	10	14	12	6	6	9	6	17	7	9	8	14	1
21	18	14	14	10	17	6	13	9	10	12	6	6	6	6	7	14	10
26	11	7	27	15	10	10	19	10	2	7	14	15	7	3	12	5	10
16	15	10	49	10	16	12	7	15	8	7	11	8	11	15	7	8	6
16	15	13	51	11	13	14	14	7	12	5	11	5	5	13	13	3	3
37	10	19	28	26	32	11	24	12	9	20	0	15	6	19	8	15	9
9	5	18	44	9	12	11	15	13	9	9	3	12	5	5	5	9	6
17	29	23	30	20	5	10	17	12	11	13	7	4	8	6	6	11	4
18	30	17	43	28	10	16	15	9	13	8	6	11	11	5	12	13	11
11	20	21	32	23	9	21	15	5	16	13	18	4	14	9	14	14	9
18	29	17	26	25	15	9	10	5	10	2	10	7	6	11	7	21	11

Kulah	Gheira	Maah	Aantah	Qehbbla	Anna	Iz	Allati	Keif	Segment Length in Tokens
8	12	6	1	7	6	8	1	2	5,016
26	9	6	5	13	6	5	5	2	5,015
29	8	1	4	5	1	4	2	1	5,023
6	16	7	1	2	4	3	8	6	5,017
16	17	5	1	3	3	7	3	7	5,016
10	13	13	1	4	6	13	5	4	5,001
20	11	4	1	6	6	1	4	4	5,021
16	11	1	12	4	4	3	2	3	5,025
8	8	3	17	7	3	2	2	1	5,049
13	9	10	4	4	2	2	4	1	5,025
16	12	7	6	3	8	2	5	1	3,857

اللسانيات الحاسوبية تنظيرا وتطبيقا : يضم الكتاب مجموعة من المباحث التي تناقش مساحة التنافذ بين اللغة والحاسوب، بوصفها حقلا بينيا، يفيد من التصورات اللسانية العامة ويوظفها في الاشتغالات الحاسوبية، لتطويع الحاسوب في خدمة اللغة . جاء الفصل الاول منه يناقش الجنبة التطبيقية للحاسوب في بعض المجالات ، نحو الترجمة، والمحادثة، والتحليل الصوتي.

ثم يوضح الفصل الثاني نقاط الالتقاء بين علمي اللغة والحاسوب ، فياتي الفصل الثالث ليوضح قيمة البرامج الصوتية الالية في التحليل الالي ، ويتكفل الفصل الرابع بتبيان طبيعة المحلل الصرفي ، والفصل الخامس فاعلية الصناعة المعجمية الالية ، واخيرا يدرس الفصل السادس التحليل الاحصائي وطريقته في تحديد الاسلوب والقصد .



اللسانيات الحاسوبية

تنظيراً وتطبيقاً

مركز الكتاب الأكاديمي
عمّان- وسط البلد- مجمع الفحيص التجاري
ص . ب : 11732 عمّان (1061) الأردن
تلفاكس: 96264619511، موبايل: 962799048009،
الموقع الإلكتروني: www.abcpub.net
A.B.Center@hotmail.com / info@abcpub.net


9789957354879

